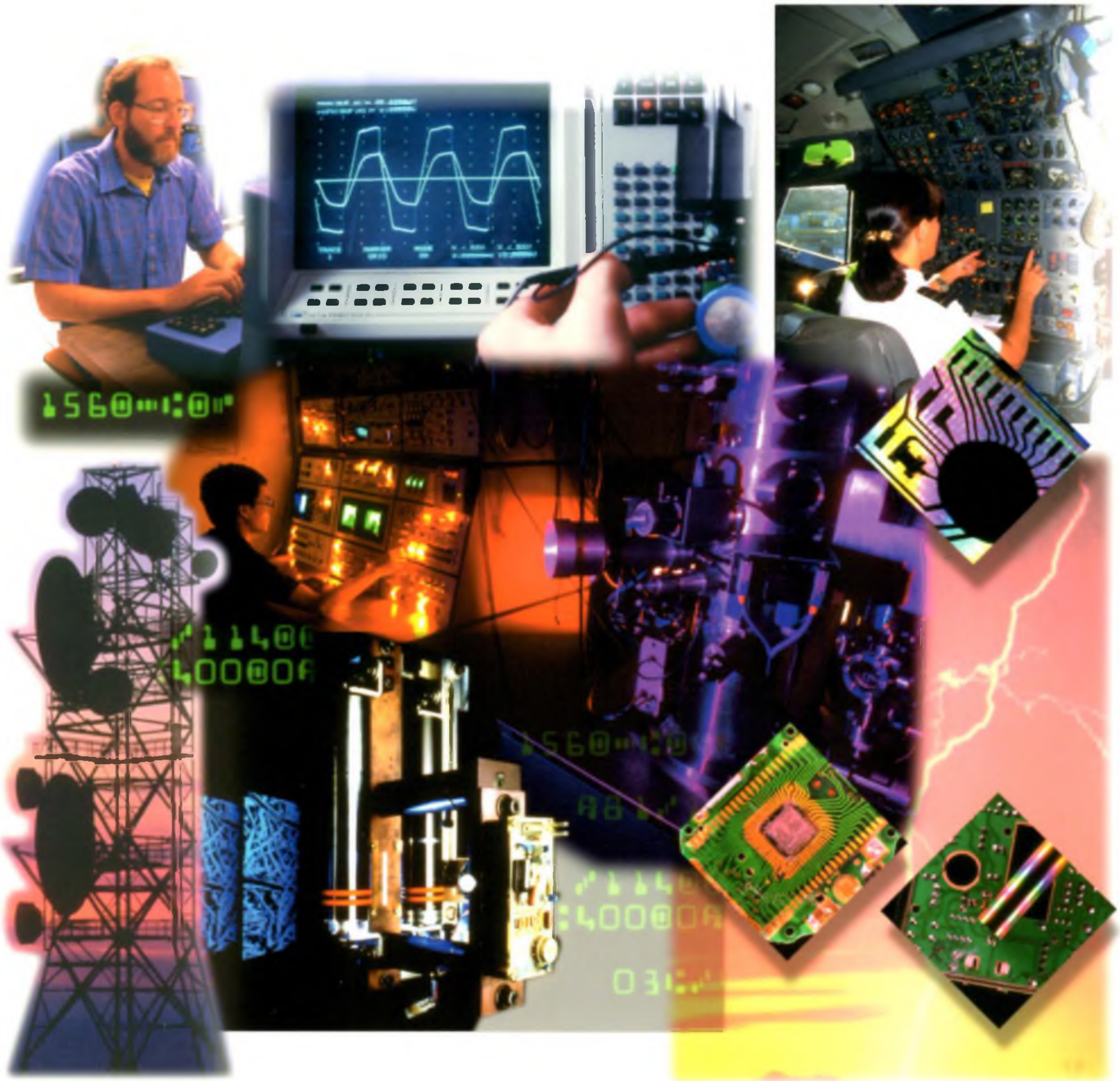
INDIVIDUAL LEARNING SYSTEM

TEXT BOOK



# SEMICONDUCTOR DEVICES

Heathkit®
EDUCATIONAL SYSTEMS

Heathkit®
EDUCATIONAL SYSTEMS
HEATHKIT COMPANY, INC., BENTON HARBOR, MI 49022
Phone: 1-800-253-0570 Fax: 1-616-925-2898

# IMPORTANT NOTICE

The SCR (#57-624) in your parts pack may be either a type 2N5061 or a type NTE5401. In any case, the 2N5061 and NTE5401 will produce the same results in your experiments.

Thank you,
Heathkit Company, Inc.

General
591-5780

# SEMICONDUCTOR DEVICES

Revised By: **Mike Sewell**
Educational Media Designer

**EE-3103B**
**595-4376-03**

**Semiconductor Devices, Third Edition**

Copyright © 1998, 1989, 1982, 1975 by Heathkit Company, Inc., Benton Harbor, Michigan 49022. All rights reserved. Printed in the United States of America. Except as permitted in the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, electronic or mechanical, including photocopying, recording, storage in a data base or retrieval system, or otherwise, without the prior written permission of the publisher.

ISBN 0-87119-210-1

# CONTENTS

Introduction . . . . . IV

Course Objectives . . . . . V

Course Outline . . . . . VI

Parts List . . . . . XIV

UNIT ONE — Semiconductor Fundamentals . . . . . 1-1

UNIT TWO — Semiconductor Diodes . . . . . 2-1

UNIT THREE — The Zener Diode . . . . . 3-1

UNIT FOUR — Semiconductor Diodes for Special Applications . . . . . 4-1

UNIT FIVE — Bipolar Transistor Operation . . . . . 5-1

UNIT SIX — Bipolar Transistor Characteristics . . . . . 6-1

UNIT SEVEN — Field Effect Transistors . . . . . 7-1

UNIT EIGHT — Thyristors . . . . . 8-1

UNIT NINE — Integrated Circuits . . . . . 9-1

UNIT TEN — Optoelectronic Devices . . . . . 10-1

INDEX . . . . . I-1

# INTRODUCTION

Semiconductor devices have revolutionized both the electronics industry and the modern world. Semiconductors, or solid-state devices, have made it possible to design circuits and equipment not possible before their development. In fact, many of the electronic devices that we now take for granted, such as hand-held calculators, digital watches, and pocket radios would not exist if not for semiconductors. The low cost of these devices make them affordable to almost everyone, resulting in a high demand that encourages further innovation.

This course describes the most common semiconductor devices and their theory of operation. Careful study of this material will lead you to a good understanding of the types and uses of solid-state components. It is to your advantage to answer all Self Test Review questions and to perform all experiments supplied with the course. You can also test your understanding of semiconductor devices by taking the Unit Examinations near the end of each unit. Most importantly, after taking the examination, review any questions that you have answered incorrectly (answers to the examination questions are supplied at the end of the unit).

Experiments included in this course complement and enhance the material covered. These experiments will make a significant contribution to your understanding of the material. Therefore, we urge you to perform these experiments to derive the maximum benefit from the course. In order to perform the experiments you will need a Volt-Ohm-Milliameter (VOM), an oscilloscope (dual channel preferred) and the parts supplied with this course.

# COURSE OBJECTIVES

When you have completed this course on semiconductor devices you will be able to:

1. Describe the electrical characteristics of materials which are classified as semiconductors.

2. Name the primary advantages that semiconductor devices have over vacuum tubes.

3. Explain how the most important semiconductor devices operate.

4. Basically describe how the most important semiconductor devices are constructed.

5. Handle semiconductor components properly without exceeding their maximum ratings or damaging them with improper handling procedures.

6. Recognize the most commonly used semiconductor packages.

7. Recognize the schematic symbols that are used to represent a wide variety of semiconductor devices.

8. Test various semiconductor devices to determine if they are functioning properly.

# COURSE OUTLINE

## UNIT 1 — SEMICONDUCTOR FUNDAMENTALS

- Introduction
- Unit Objectives
- Unit Activity Guide
- The Importance of Semiconductors
  - Advantages
  - Disadvantages
- Semiconductor Materials
- Germanium and Silicon Atoms and Crystals
  - Semiconductor Atoms
  - Semiconductor Crystals
- Conduction in Intrinsic Germanium and Silicon
  - Low Temperature Characteristics
  - High Temperature Characteristics
  - Holes
  - Current Flow
- Conduction in Doped Germanium and Silicon
  - N-Type Semiconductors
  - P-Type Semiconductors
- Unit Summary
- Unit Examination

## UNIT 2 — SEMICONDUCTOR DIODES

- Introduction
- Unit Objectives
- Unit Activity Guide
- The PN Junction
  - Ions
  - Junction Diodes
    - Depletion Region
    - Barrier Voltage
- Diode Biasing
  - Forward Bias
  - Reverse Bias
- Diode Characteristics
  - Germanium Diode
    - Forward Characteristics
    - Reverse Characteristics
  - Silicon Diode
    - Forward Characteristics
    - Reverse Characteristics
  - Diode Ratings
  - Temperature Considerations
  - Diode Symbols
- Experiment 1 — Semiconductor Diode Characteristics
- Diode Construction
  - Grown Junctions
  - Alloyed Junctions
  - Diffused Junctions
  - Diode Packaging
- Experiment 2 — Testing Semiconductor Diodes
- Diode Applications
  - Rectification
  - Other Applications for Diodes
- Unit Summary
- Unit Examination

## UNIT 3 — THE ZENER DIODE

- Introduction
- Unit Objectives
- Unit Activity Guide
- Zener Diode Characteristics
  - Voltage-Current Characteristics
  - Zener Voltage
- Power Dissipation in Zener Diodes
  - Power-Temperature Curves
  - Derating Factor
- Current Limitations in Zener Diodes
- Effects of Temperature on Zener Voltage
- Temperature Compensated Zener Diodes
- Zener Diode Impedance
- Zener Diode Packages
- Voltage Regulation with Zener Diodes
  - The Basic Zener Diode Regulator
  - Designing a Zener Regulator Circuit
- Experiment 3 — Zener Diodes
- Unit Summary
- Unit Examination

## UNIT 4 — SEMICONDUCTOR DIODES FOR SPECIAL APPLICATIONS

- Introduction
- Unit Objectives
- Unit Activity Guide
- The Tunnel Diode
  - Voltage-Current Characteristics
  - What is Negative Resistance?
  - Construction
  - Applications
- The Varactor Diode
  - Electrical Characteristics
  - Operating Efficiency
  - Construction
  - Applications
- Experiment 4 — The Varactor Diode
- High Frequency Diodes
  - PIN Diodes
  - Schottky Diodes
  - Gunn Effect Devices
  - IMPATT Diodes
- Unit Summary
- Unit Examination

## UNIT 5 — BIPOLAR TRANSISTOR OPERATION

Introduction
Unit Objectives
Unit Activity Guide
PNP and NPN Configurations
Transistor Construction
    Construction Techniques
    Packaging
Basic Transistor Action
    Biasing NPN Transistors
    Biasing PNP Transistors
Transistor Amplification
Transistor Circuit Arrangements
    Common Base Circuits
    Common Emitter Circuits
    Common Collector Circuits
Testing Bipolar Transistors
    Test Procedures
    Testing NPN Transistors
    Testing PNP Transistors
    Identifying a Transistor
Experiment 5 — Testing Bipolar Transistors
Experiment 6 — Transistor Characteristics
Unit Summary
Unit Examination

## UNIT 6 — BIPOLAR TRANSISTOR CHARACTERISTICS

- Introduction
- Unit Objectives
- Unit Activity Guide
- Characteristics of Common-Base Circuits
  - Collector Characteristic Curves
  - Current Gain
  - Alpha Cutoff Frequency
  - Collector-To-Base Leakage Current
- Characteristics of Common-Emitter Circuits
  - Collector Characteristic Curves
  - Current Gain
  - Beta Cutoff Frequency
  - Collector-To-Emitter Leakage Current
- Experiment 7 — Common-Emitter Amplifier Characteristics — I
- Characteristics of Common-Collector Circuits
  - Input Resistance
  - Output Resistance
- Maximum Transistor Ratings
  - Collector Breakdown Voltage
  - Emitter Breakdown Voltage
  - Maximum Collector and Emitter Currents
  - Maximum Collector Dissipation
  - Temperature Ratings
  - Current Gain-Bandwidth Product
- Experiment 7A — Common Base Amplifier Characteristics
- Experiment 7B — Common Emitter Amplifier Characteristics — II
- Experiment 7C — Common Collector Amplifier Characteristics
- Unit Summary
- Unit Examination

## UNIT 7 — FIELD EFFECT TRANSISTORS

- Introduction
- Unit Objectives
- Unit Activity Guide
- The Junction FET
  - Construction of a JFET
  - Operation of a JFET
  - Gate-To-Source Cutoff Voltage
  - Pinch-Off Voltage
  - Drain Characteristic Curves
    - Transconductance
    - Symbols
- Experiment 8 — Junction FET Characteristics
- The Insulated Gate FET
  - Depletion-Mode Devices
  - Enhancement-Mode Devices
  - CMOS
  - Safety Precautions
- FET Circuit Arrangements
  - Common-Source Circuits
  - Common-Gate Circuits
  - Common-Drain Circuits
- Unit Summary
- Unit Examination

## UNIT 8 — THYRISTORS

- Introduction
- Unit Objectives
- Unit Activity Guide
- Silicon Controlled Rectifiers
  - Basic Construction and Operation
  - V-I Characteristics
  - Practical Applications
- Experiment 9 — SCR Characteristics
- Bidirectional Triode Thyristors
  - Basic Construction and Operation
  - V-I Characteristics
  - Applications
- Bidirectional Trigger Diodes
  - Basic Construction and Operation
  - V-I Characteristics
  - Applications
- Unijuction Transistors
  - Basic Construction
  - Operation
  - V-I Characteristics
  - Applications of the UJT
  - Programmable UJT's
- Experiment 10 — UJT Characteristics
- Unit Summary
- Unit Examination

## UNIT 9 — INTEGRATED CIRCUITS

- Introduction
- Unit Objectives
- Unit Activity Guide
- The Importance of Integrated Circuits
  - Advantages
  - Disadvantages
- IC Construction
  - Monolithic IC's
    - Bipolar IC's
    - MOS IC's
  - Thin-Film IC's
  - Thick-Film IC's
  - Hybrid IC's
  - IC Packages
- Applications of the IC's
  - Digital IC's
  - Linear IC's
- Experiment 11 — Operational Amplifier Characteristics
- Unit Summary
- Unit Examination

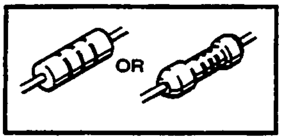
## UNIT 10 — OPTOELECTRONIC DEVICES

- Introduction
- Unit Objectives
- Unit Activity Guide
- Basic Principles of Light
    - Characteristics
    - Units of Measurement
        - Radiometric System
        - Photometric System
- Light Sensitive Devices
    - Photoconductive Cells
    - .Photovoltaic Cells
    - Photodiodes
    - Phototransistors
- Experiment 12 — Phototransistor Characteristics
- Light-Emitting Devices
    - LED Operation
    - LED Construction
    - LED Characteristics
    - LED Applications
- Liquid Crystal Displays
- Experiment 13 — LED Characteristics
- Unit Summary
- Unit Examination

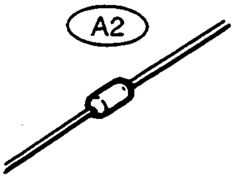
# PARTS LIST

This parts list contains all of the parts used in the experiments which you will perform with this course. The key number in the parts list corresponds to the numbers in the parts pictorial. Some parts are packaged in envelopes. Except for this initial parts check, keep these parts in their envelopes until they are called for in the experiment. A container is provided so that you can keep the small parts together in one place.

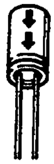
A1



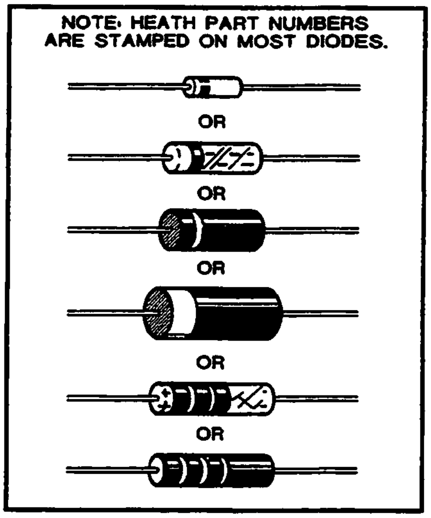
A2



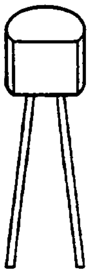
A3



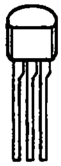
A4



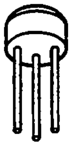
A5



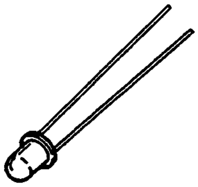
A6



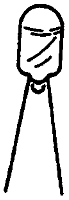
A7



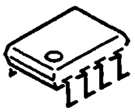
A8



A9



A10



| KEY No. | PART No. | QTY. | DESCRIPTION |
|---|---|---|---|
| **RESISTORS (1/2-watt ± 5% tolerance unless otherwise stated)** | | | |
| A1 | 6-470 | 1 | 47 Ω (yellow, violet, black) |
| A1 | 6-101 | 2 | 100 Ω (brown, black, brown) |
| A1 | 6-471 | 1 | 470 Ω (yellow, violet, brown) |
| A1 | 6-102 | 2 | 1000 Ω (brown, black, red) |
| A1 | 6-472 | 1 | 4700 Ω (yellow, violet, red) |
| A1 | 6-103 | 2 | 10 KΩ (brown, black, orange) |
| A1 | 6-104 | 1 | 100 KΩ (brown, black, yellow) |
| A1 | 6-334 | 1 | 330 KΩ (orange, orange, yellow) |
| A1 | 6-222 | 1 | 2200 Ω (red, red, red) |
| A1 | 6-221 | 1 | 220 Ω (red, red, brown) |
| A1 | 6-101-2 | 1 | 100 Ω (brown, black, brown) 2-watt |
| **CAPACITORS** | | | |
| A2 | 21-786 | 1 | 0.1 μF ceramic disc capacitor |
| A3 | 25-880 | 1 | 10 μF electrolytic capacitor |
| **DIODES** | | | |
| A4 | 56-161 | 1 | 1N5231B zener diode (5.1 V) |
| A4 | 56-28 | 1 | 1N4148 silicon diode |
| A4 | 56-602 | 1 | Germanium diode (selected) |
| A5 | 56-674 | 1 | MV209 varactor diode |
| **TRANSISTORS** | | | |
| A6 | 417-201 | 1 | X29A829 PNP transistor |
| A6 | 57-624 | 1 | SCR 2N5061 |
| A6 | 417-183 | 1 | 2N4894 unijunction transistor |
| A6 | 417-291 | 1 | JFET N-channel, 2N5458 |
| A6 | 417-801 | 1 | MPSA20 transistor NPN |
| A7 | 417-851 | 1 | Phototransistor FPT 100 |
| **MISCELLANEOUS** | | | |
| A8 | 412-640 | 1 | LED (light-emitting diode) |
| A9 | 412-601 | 1 | Lamp (2174 D or equiv.) |
| A10 | 442-22 | 1 | 741 Operational Amplifier |
| | 344-59 | 10' | Hookup wire |
| | 266-962 | 1 | Small parts container |

## *Unit 1*

# SEMICONDUCTOR FUNDAMENTALS

# CONTENTS

Introduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-3

Unit Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-3

Unit Activity Guide. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-4

The Importance of Semiconductors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-5

Advantages . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-5

Disadvantages . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-7

Semiconductor Materials . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-10

Germanium and Silicon Atoms and Crystals . . . . . . . . . . . . . . . . . . . . . .Page 1-12

Semiconductor Atoms. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-12

Semiconductor Crystals. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-14

Conduction in Intrinsic Germanium and Silicon. . . . . . . . . . . . . . . . . . . .Page 1-17

Low Temperature characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-17

High Temperature characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-17

Holes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-18

Current Flow . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-18

Conduction in Doped Germanium and Silicon . . . . . . . . . . . . . . . . . . . .Page 1-21

N-Type Semiconductors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-21

P-Type Semiconductors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-22

Unit Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-25

Unit Examination . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-26

Examination Answers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 1-29

# INTRODUCTION

In this unit on semiconductor fundamentals, you will learn how and why semiconductors are used in electronics. You will examine the various types of semiconductor materials and study their electrical characteristics. The information presented in this unit is extremely important because it forms the basic foundation upon which you will eventually build your knowledge of semiconductor devices. Examine the Unit Objectives closely so that you will know exactly what you will achieve with this unit. Then carefully follow the instructions contained in the Unit Activity Guide to be sure that you perform all of the necessary steps to complete this unit successfully. As you perform each step in the Unit Activity Guide check it off in the corresponding box so that you can keep track of your work as you progress. Spaces are provided so that you can conveniently keep track of the time you spend on each activity.

# UNIT OBJECTIVES

When you have completed this unit, you should be able to:

1. Identify three common semiconductor devices.
2. Describe some of the uses of semiconductor devices in electronic equipment.
3. List at least five advantages semiconductor devices have over other components having similar capabilities.
4. Identify the two most commonly used semiconductor materials.
5. Describe the crystal lattice structure of semiconductors
6. Define the electrical characteristics of semiconductors.
7. Describe the difference between intrinsic and doped semiconductors.
8. Define the term "hole" as applied to semiconductors.
9. Recognize the difference between majority carriers and minority carriers in a doped semiconductor.

# UNIT ACTIVITY GUIDE

**Completion Time**

☐ Read "The Importance of Semiconductors." ____________

☐ Answer Self Test Review Questions 1-12. ____________

☐ Read "Semiconductor Materials." ____________

☐ Answer Self Test Review Questions 13-16. ____________

☐ Read "Germanium and Silicon Atoms and Crystals." ____________

☐ Answer Self Test Review Questions 17-28. ____________

☐ Read "Conduction in Intrinsic Germanium and Silicon." ____________

☐ Answer Self Test Review Questions 29-38. ____________

☐ Read "Conduction in Doped Germanium and Silicon." ____________

☐ Answer Self Test Review Questions 39-47. ____________

☐ Read the Unit Summary. ____________

☐ Complete the Unit Final Examination. ____________

☐ Check the Examination Answers. ____________

# THE IMPORTANCE OF SEMICONDUCTORS

Semiconductors essentially serve as the basic building materials which are used to construct some very important electronic components. These semiconductor components are in turn used to construct electronic circuits and equipment. The three most commonly used semiconductor devices are diodes, transistors, and integrated circuits. However, other special components are also available.

The primary function of semiconductor devices in electronic equipment is to control currents or voltages in such a way as to produce a desired end result. For example, diodes can be used as rectifiers to produce pulsating dc from ac. A transistor can be used as a variable resistance to control the current in a heating element. Or, an integrated circuit can be used to amplify and demodulate a radio signal. All of these components are made of specially processed semiconductor material such as silicon and germanium.

Semiconductor devices are extremely small, lightweight components which consume only a small amount of power and are highly efficient and reliable. The vacuum tubes that were once widely used in practically all types of electronic equipment have been virtually replaced by the newer and better semiconductor devices. Let's consider some of the specific reasons for this significant transition from the use of vacuum tubes to semiconductor components in electronic equipment.

## Advantages

Components which are made of semiconductor materials are often referred to as *solid-state* components because they are made from solid materials. Because of this solid-state construction, these components are more rugged than vacuum tubes which are made of glass, metal, and ceramic materials. Because of this ruggedness, semiconductor devices are able to operate under extremely hazardous environmental conditions. This ruggedness is responsible for the reliability of solid-state devices.

The solid-state construction also eliminates the need for filaments or heaters as found in all vacuum tubes. This means that additional power is not required to operate the filaments and component operation is cooler and more efficient. By eliminating the filaments, a prime source of trouble is also avoided because the filaments generally have a limited life expectancy. The absence of filaments also means that a warm-up period is not required before the device can operate properly. In other words, the solid-state component operates the instant it receives electrical power.

Solid state components are also able to operate with very low voltages (between 1 and 25 volts) while vacuum tubes usually require an operating voltage of 100 volts or more. This means that solid state components generally use less power than vacuum tubes and are, therefore, more suitable for use in portable equipment which obtains its power from batteries. The lower voltages are also much safer to work with. Pocket-size radios, hand held calculators, and small battery operated television receivers are typical examples of devices which take advantage of highly efficient, power saving components.

The small size of the solid-state component also makes it suitable for use in portable electronic equipment. Although equipment of this type can be constructed with vacuum tubes, such equipment would be much larger and heavier. A typical transistor is only a fraction of an inch high and wide while a vacuum tube of comparable performance may be an inch or more wide and several inches high. The small size also means a significant weight savings.

Solid-state components are much less expensive than comparable vacuum tube components. The very nature of a solid-state component makes it suitable for production in mass quantities which brings about a high cost saving. In fact, a large number of solid-state components can be constructed as easily and quickly as a single component.

The most sophisticated semiconductor devices are integrated circuits. These are complete circuits where all of the components are constructed with semiconductor materials in a single microminiature package. These devices not only replace individual electronic circuits but also complete pieces of equipment or entire systems. Entire computers and radio receivers can be constructed as a single device no larger than a typical transistor. Integrated circuits have taken us one step farther in improving electronic equipment through the use of semiconductor materials. All electronic equipment has benefited from solid state components and particularly from the development of integrated circuits.

# Disadvantages

Although solid-state components have many advantages over the vacuum tubes that were once widely used, they also have several inherent disadvantages. First, solid-state components are highly susceptible to changes in temperature and can be damaged if they are operated at extremely high temperatures. Additional components are often required simply for the purpose of stabilizing solid-state circuits so that they will operate over a wide temperature range. Solid-state components may be easily damaged by exceeding their power dissipation limits and they may also be occasionally, damaged when their normal operating voltages are reversed. In comparison, vacuum tube components are not nearly as sensitive to temperature changes or improper operating voltages.

There are still a few areas where semiconductor devices cannot replace tubes. This is particularly true in high power, ultra high radio frequency applications. However, as semiconductor technology develops, these limitations are gradually being overcome.

Despite the several disadvantages just mentioned, solid-state components are still the most efficient and reliable devices to be found. They are used in all new equipment designs and new applications are constantly being found for these devices in the military, industrial, and consumer fields. The continued use of semiconductor materials to construct new and better solid-state components is almost assured because the techniques used are constantly being refined thus making it possible to obtain even superior components at less cost.

Semiconductors have had a profound effect on the design and application of electronic equipment. Not only have they greatly improved existing equipment and techniques by making them better and cheaper, but also they have permitted us to do things that were not previously possible. Semiconductors have revolutionized the electronic industry and they continue to show their greater potential. Your work in electronics will always involve semiconductor devices. Examine this unit on semiconductor fundamentals carefully and you will benefit from the resulting knowledge.

# Self-Test Review

1. Components which are made from semiconductor materials are often referred to as ______________ components.

2. Components made from semiconductor materials do not require a warm-up period.

   a. True
   b. False

3. Transistors are more reliable and have a longer life expectancy than tubes.

   a. True
   b. False

4. Components made from semiconductors require high operating voltages.

   a. True
   b. False

5. Practically all modern portable and compact electronic equipment utilizes ______________ components.

6. Semiconductor type components have almost completely replaced the older __________ __________ components.

7. Components made from semiconductors are highly susceptible to temperature changes.

   a. True
   b. False

8. By using semiconductor materials it is possible to construct entire circuits.

   a. True
   b. False

9. List four advantages of semiconductor components over vacuum tubes.

   a. ________________ b. ________________
   c. ________________ d. ________________

10. The three most common semiconductor components are ______________, ______________ and ______________.

11. The primary function of a semiconductor component is to ____________ the ______________ in a circuit.

12. The most advanced, complex and sophisticated solid-state device is the ______________ ______________.

# Answers

1. solid-state
2. a. True
3. a. True
4. b. False
5. semiconductor or solid-state
6. vacuum tube
7. a. True
8. a. True (integrated circuits)
9. a. lower cost
   b. less power consumption
   c. smaller, lighter
   d. more rugged and reliable
10. diode, transistor and integrated circuit
11. control the current (or voltage)
12. integrated circuit

# SEMICONDUCTOR MATERIALS

The term semiconductor is used to describe any material that has characteristics which fall between those of insulators and conductors. In other words, a semiconductor will not pass current as readily as a conductor not will it block current as effectively as an insulator. Some semiconductor materials are actually pure elements which are found in the periodic table of elements while other semiconductors may be classified as compounds. Typical examples of semiconductor materials that are natural elements are germanium (Ge) and silicon (Si).

The semiconductors that are suited to the greatest variety of electronic applications are germanium and silicon. Germanium is a brittle, grayish-white element that was discovered in 1886. This material may be recovered from the ash of certain types of coals in the form of germanium dioxide powder. This powder may then be reduced to pure germanium which is in solid form.

Silicon is a non-metallic element which was discovered in 1823. This material is found extensively in the earth's crust. A white or sometimes colorless compound known as silicon dioxide (also called silica) occurs abundantly in forms such as sand, quartz, agate, and flint. These silicon compounds can be chemically reduced to obtain pure silicon which is in a solid form. These two materials have atomic structures which may be easily altered to obtain specific electrical characteristics.

Once the pure material is available, it must then be suitably modified to give it the qualities necessary to construct a semiconductor device for a definite application.

# Self-Test Review

13. Materials which have characteristics that fall midway between those of insulators and conductors are called ____________.

14. The two semiconductor materials most commonly used to manufacture electronic components are ______________ and ______________.

15. The resistance of a semiconductor material compared to that of a good conductor like copper is

    a. higher
    b. lower
    c. about the same

16. Semiconductor devices are made directly from pure semiconductor materials which have not been altered in any way.

    a. True
    b. False

## Answers

13. semiconductors

14. silicon and germanium

15. (a) higher — The resistance of a semiconductor is higher than that of a conductor but lower than that of an insulator.

16. (b) False

# GERMANIUM AND SILICON ATOMS AND CRYSTALS

From your previous studies you know that metals such as copper and aluminum are used to carry current in an electrical circuit. You learned that these metals are classified as conductors because they offer minimum opposition to current flow. In previous studies you also learned that materials such as glass, rubber, and ceramic oppose the flow of electrical current and are therefore classified as insulators. You will now examine the atomic structure of two materials (germanium and silicon) which have characteristics that are between those of conductors and insulators and find out why these materials fall into a third classification know as semiconductors.

## Definitions

| | |
|---|---|
| Elements: | One of the known chemical materials that cannot be subdivided into simpler substances. |
| Atoms: | The smallest portion of an element that still exhibits all the characteristics of that element. |

## Semiconductor Atoms

Before we actually examine the structure of germanium and silicon atoms and crystals we must consider some important rules which pertain to the number and placement of the electrons which revolve around the nucleus of all atoms. Atoms contain three basic components: protons, neutrons, and electrons. The protons and neutrons are located in the nucleus or center of the atom while the electrons revolve around the nucleus in orbits. The atom of each particular element will have a specific number of protons in its nucleus and an equal number of electrons in orbit if the atom is neutral (has no charge). However, the exact manner in which the electrons are arranged around the nucleus is extremely important in determining the electrical characteristics of the element. Generally, each electron has its own orbit, but certain orbits are grouped together to produce what is referred to as a shell. For all of the elements that are known to exist, there can only be a maximum of seven shells.

The shell nearest the nucleus can only hold 2 electrons while the second shell can hold a maximum of 8 electrons. The third shell cannot hold more than 18 electrons and the fourth can hold no more than 32 and so on. The outermost shell of a particular atom is called the valence shell and the electrons that orbit within this shell are referred to as valence electrons.

The arrangement of the protons and electrons in three different atoms are shown in Figure 1-1. Notice that the hydrogen atom has only one shell, while the carbon atom has two and the copper atom has four. Also notice that some of the shells contain less than the maximum number of electrons allowed.

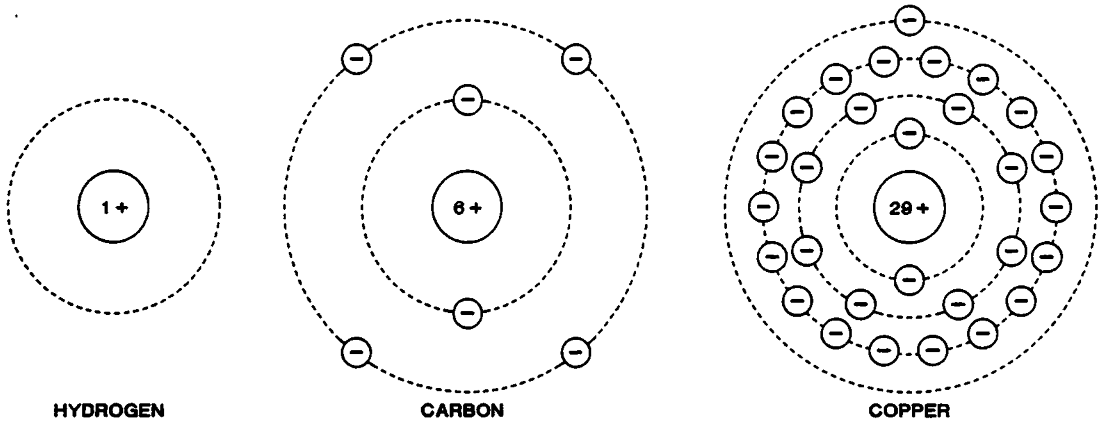


Figure 1-1
Diagram of typical Atoms.

In any particular atom, the outer shell can never hold more than 8 electrons. When exactly 8 electrons are present in the outer shell, the atom is considered to be completely stable and it will neither give up or accept electrons easily. Elements which have atoms of this type are neon and argon. These elements are classified as inert gases and they resist any sort of electrical or chemical activity.

When and atom has 5 or more electrons in its outer shell, it tries to fill its shell so that it can reach a stable condition. Elements of this type make good insulators because the individual atoms try to acquire electrons instead of giving them up. Therefore, the free movement of electrons from one atom to the next is inhibited. When an atom has less than 4 electrons in its outer shell it tends to give up these outer electrons easily. Elements which have atoms of this type make good conductors because they contain a large number of free-moving electrons which can randomly drift from one atom to the next.

When an atom contains exactly 4 electrons in its outer shell, it does not readily give up or accept electrons. Elements which contain atoms of this type do not make good insulators or conductors and are, therefore, referred to as semiconductors. The element carbon is a typical example of a semiconductor material. Notice that the carbon atom shown in Figure 1-1 has exactly 4 electrons in its outermost shell.

The two semiconductor materials most commonly used in the construction of transistors and other types of related components are germanium and silicon. Both of these materials are made up of atoms which have 4 electrons in their outermost or valence shells. A single germanium atom is shown in Figure 1-2. Notice that the germanium atom has four shells and the distribution of the electrons from the first shell to the outer shell is 2, 8, 18, and 4. Therefore, a total of 32 electrons rotate around the nucleus of the atom and a total of 32 protons are contained within the nucleus.

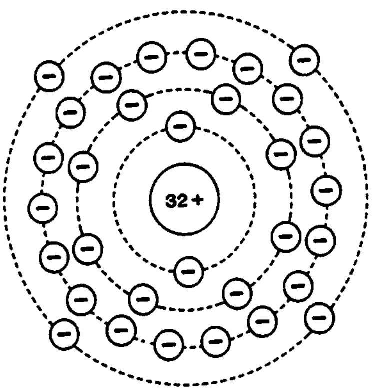


Figure 1-2
The Germanium Atom.

A single silicon atom is shown in Figure 1-3. Notice that this atom has only three shells and the distribution of the electrons from the first shell to the valence shell is 2, 8 and 4. The atom has a total of 14 electrons revolving around its nucleus and a total of 14 protons in its nucleus.

The important point to consider when examining the structure of germanium and silicon atoms is that both of these atoms have exactly 4 valence electrons. Although both of these atoms contain additional electrons, it is the valence electrons that determine the relative ease with which electrical current can flow through the germanium and silicon materials. To simplify the discussions and illustrations which follow, we will use diagrams of the germanium and silicon atoms which show only the four valence electrons surrounding a central core which consists of the nucleus and the inner shells. These simplified diagrams are shown in Figure 1-4. Notice that the core of the germanium atom is identified by the symbol Ge and the core of the silicon atom is identified by the symbol Si. In each case the nucleus is surrounded by four electrons.

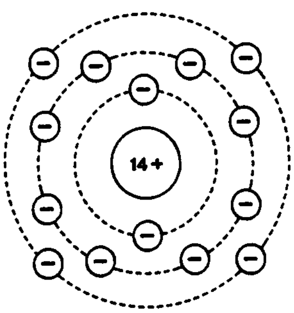


Figure 1-3
The Silicon Atom.

## Semiconductor Crystals

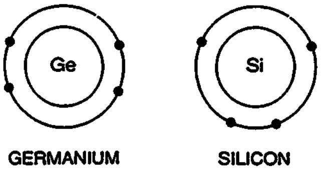


Figure 1-4
Simplified Germanium and Silicon Atoms.

The individual atoms within a semiconductor material such as germanium are arranged as shown in Figure 1-5. Each atom shares its four valence electrons with four neighboring atoms as shown. This sharing of electrons creates a bond which holds the atoms together. This electron-pair bond is commonly referred to as a *covalent bond*, and it occurs because each of the atoms in the structure tries to take on additional electrons in order to fill its valence shell with eight electrons. The end result is a structure which has a lattice-like appearance and is often referred to as a crystal lattice.

Silicon atoms combine in the same manner as germanium atoms to form the same type of crystaline structure shown in Figure 1-5. The silicon atoms maintain covalent bonds just like the germanium atoms.

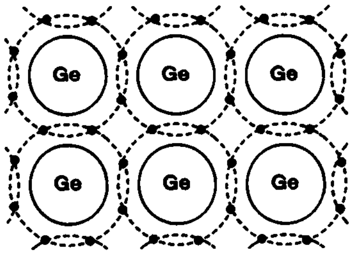


Figure 1-5
Simplified Diagram of Germanium Crystal Structure.

The germanium and silicon crystal lattices just described are free from impurities and therefore represent pure or ideal materials. Such crystals are often referred to as *intrinsic* materials. The construction of solid-state components such as transistors depends on the use of these pure or intrinsic semiconductor materials.

## Self Test Review

17. A neutral atom will have an equal number of ______________ and ______________.

18. The electrons surrounding the nucleus of an atom are grouped into various ______________.

19. There can never be more than _________ shells in any atom.

20. The shell closest to the nucleus of an atom cannot hold more than ______ electrons.

21. The outermost shell is referred to as the ______________ shell.

22. The outermost shell can never hold more than _______ electrons.

23. A neon atom is completely stable because it has ______________ electrons in its outer shell.

24. The electrical characteristics of an element is determined by number of electrons in the ______________ shell of its atoms.

25. Elements which are made up of atoms that have 4 electrons in their outer shells are called ______________.

26. The atoms in a semiconductor are arranged in the form of a __________ ______________.

27. Each atom in a semiconductor material shares its electrons with four neighboring atoms to form ______________ ______________.

28. Intrinsic semiconductors do not contain ______________.

## Answers

17. protons and electrons
18. shells
19. 7
20. 2
21. valence
22. 8
23. 8
24. outer or valence
25. semiconductors
26. crystal lattice
27. covalent bonds
28. impurities

# CONDUCTION IN INTRINSIC GERMANIUM AND SILICON

Due to the crystalline structure of pure semiconductor materials such as germanium and silicon, each nucleus within the material sees eight valence electrons even though each atom actually has only four. Therefore, each atom tends to be stable and will not easily give up or accept electrons. However, this does not mean that pure semiconductors must, under all conditions, resist any sort of electrical activity in the same manner as the inert gases previously described. The reason for this is that another factor must be considered. This factor is temperature. The electrical characteristics of semiconductor materials is highly dependent upon temperature.

## Low Temperature Characteristics

At extremely low temperatures, the valence electrons are held tightly to their parent atoms, which are in covalent bonds, and are not allowed to drift through the crystalline structures of either semiconductor material. Since the valence electrons cannot drift from one atom to the next, the material cannot support current flow at this time. Therefore, at extremely low temperatures, pure germanium and silicon crystals function as insulators.

## High Temperature Characteristics

As the temperature of a germanium or silicon crystal is increased, the valence electrons within the material become agitated and some of them will occasionally break away from the covalent bonds. Therefore, a small number of electrons will be free to drift from one atom to the next in a random manner. These free-moving electrons or *free electrons* are able to support a small amount of electrical current if a voltage is applied to the semiconductor material. In other words, as the temperature of the semiconductor material increases, the material begins to acquire the characteristics of a conductor. For all practical purposes, however, enough heat energy is available even at room temperature to produce a small number of free electrons that can support a small amount of current. Only when the semiconductor materials are exposed to extremely high temperatures, can a point be reached where they will conduct current as well as an ordinary conductor. Under normal conditions, however, this high temperature usage of semiconductors is never encountered.

# Holes

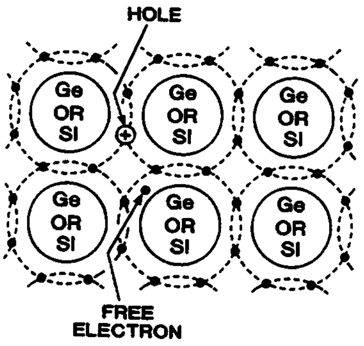


Figure 1-6
An Electron-Hole Pair
in a Semiconductor Material

To understand exactly why a semiconductor is able to allow current to flow, we must take a closer look at the internal structure of the material. When an electron breaks away from a covalent bond, an open space or vacancy exists in the bond. The space that was previously occupied by the electron is generally referred to as a *hole*. A hole simply represents the absence of an electron. Since an electron has a negative charge, the hole represents the absence of a negative charge. This means that the hole has the characteristics of a positively charged particle. Each time an electron breaks away from a covalent bond, a hole is created. Each corresponding electron and hole is referred to as an electron-hole pair. A typical electron-hole pair is shown in Figure 1-6. Notice that the hole is represented by a plus sign to indicate it is positive but the electron is simply shown as a dot although it does have a negative charge. The semiconductor material shown could be either germanium or silicon as indicated.

The number of electron-hole pairs produced within a semiconductor material will increase with temperature. However, even at room temperature, a small number of electron-hole pairs will exist. Some of the free electrons will tend to drift randomly, however, the holes will try to absorb some of the electrons. This means that some electrons will simply jump from one shell to another shell which contains a hole. If an electron jumps from one shell to fill in a hole, another hole is created when the electron leaves the shell. The hole, therefore, appears to move in the opposite direction of the electron. If another electron moves into the hole that was just created, another hole is produced and the previous hole appears to move randomly through a pure semiconductor material.

# Current Flow

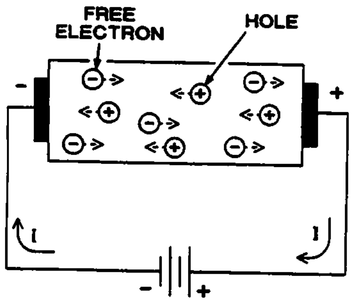


Figure 1-7
Current flow in a
Semiconductor Material

When a pure semiconductor material such as germanium or silicon is subjected to a difference-of-potential or voltage as shown in Figure 1-7, the negatively charged free electrons are attracted to the positive terminal of the voltage source. The positive holes that are created by the free electrons drift toward the negative terminal of the voltage source. As the free electrons flow into the positive terminal of the voltage source, an equal number of electron leave the negative terminal of the voltage source. These electrons are injected into the left side of the semiconductor material as shown where many of these electrons are captured or absorbed by holes. As the holes and electrons recombine in this manner, the holes cease to exist. Therefore the holes constantly drift to the left and then disappear while the electrons flow to the right side of the material where they are then drawn out of the material and into the positive terminal of the voltage source.

It is important to remember that current flow in a semiconductor material consists of both electrons and holes. The holes function like positively charged particles while the electrons are actually negatively charged particles. The holes and electrons flow in opposite directions and the number of electron-hole pairs produced within a material increase as the temperature of the material increases. Since the amount of current flowing in a semiconductor is determined by the number of electron-hole pairs in the material, the ability of a semiconductor material to pass current increases as the temperature of the material increases.

It is also important to note that current flow in a semiconductor is somewhat different than current flow in a conductor. When we consider current flow in a semiconductor, we must consider the movement of holes as well as electrons. However, in a conductor we are concerned only with the number of free-electrons that are available.

## Self Test Review

29. At extremely low temperatures, a semiconductor is basically an ____________.

30. At extremely high temperatures, a semiconductor will function as a ____________.

31. Electrons which break away from covalent bonds and drift through a semiconductor are called ____________ ____________.

32. When an electron breaks away from a covalent bond, a hole is created.
    a. True
    b. False

33. An electron has a ____________ charge.

34. A hole has a ____________ charge.

35. When a pure semiconductor is subjected to a voltage, free electrons flow toward the ____________ terminal.

36. Holes move in the same direction as the electrons.
    a. True
    b. False

37. The number of electron-hole pairs within a semiconductor increases as temperature ____________.

38. The resistance of a semiconductor ____________ as the temperature decreases.

## Answers

29. insulator

30. conductor

31. free electrons

32. a. True

33. negative

34. positive

35. positive

36. b. False — Holes move in the opposite direction of electrons.

37. increases

38. increases — Decreasing the temperature reduces the number of electron-hole pairs, so current flow is restricted. In other words, the electrical resistance increases.

# CONDUCTION IN DOPED GERMANIUM AND SILICON

Pure semiconductor materials contain only a small number of electrons and holes at room temperature and therefore conduct very little current. However, the conductivity of these materials can be increased considerably by a process known as doping. Pure semiconductor materials such as germanium and silicon are doped by adding other materials, called impurities, to them when they are produced. Basically there are two types of impurities that are added to germanium and silicon crystals. One type of impurity is referred to as a pentavalent material because it is made up of atoms which have five valence electrons. The second type of impurity is referred to as a trivalent material because each of the atoms in this material has three valence electrons.

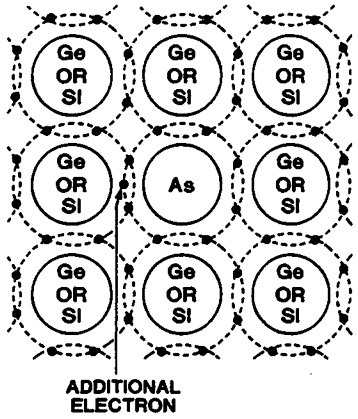


Figure 1-8
Semiconductor Material Doped with Arsenic.

## N-Type Semiconductors

When a pure semiconductor material is doped with a pentavalent element, such as arsenic (As), some of the atoms in the crystal lattice structure of the tetravalent semiconductor are replaced by arsenic atoms. As a result, the crystal lattice of the semiconductor is like that shown in Figure 1-8. As shown in the figure, the arsenic atom has replaced one of the semiconductor atoms and is sharing four of its valence electrons with adjacent atoms in a covalent bond. However, the fifth electron is not part of a covalent bond and can be easily freed from the atom. This arsenic atom is called a donor atom because it donates a free electron to the crystal lattice. Actually, there is a large number of donor atoms in the crystal lattice. Consequently, there are many free electrons in the semiconductor.

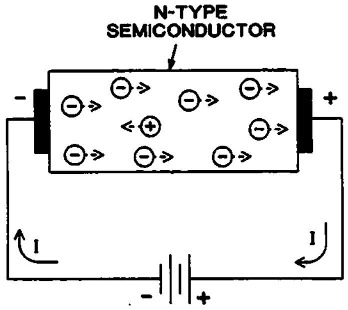


Figure 1-9
Conduction in an N-Type Semiconductor.

If a voltage is applied to an N-type semiconductor as shown in Figure 1-9, the free electrons contributed by the donor atoms will flow toward the positive terminal of the battery. However, some additional free electrons will also flow toward the positive terminal. These additional free electrons are produced as electrons break away from their covalent bonds to create electron - hole pairs and is identical to the action which takes place in a pure semiconductor material. The corresponding holes which are produced are then moved toward the negative terminal.

At normal room temperature the number of free electrons provided by the donor atoms will greatly exceed the number of holes and electrons that are produced by the breaking of covalent bonds. This means that the number of electrons flowing in the N-type semiconductor will greatly exceed the number of holes. The electrons, being in the majority, are therefore referred to as the majority carriers while the holes, which are in the minority, are referred to as the minority carriers.

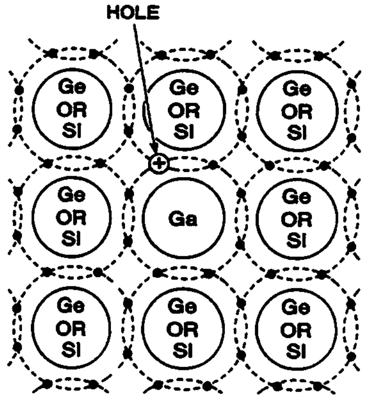


Figure 1-10
Semiconductor Material
Doped with Gallium.

## P-Type Semiconductors

Doping pure semiconductor material with a trivalent element, such as gallium (Ga), causes some of the semiconductor tetravalent atoms to be displaced by trivalent atoms. As shown in Figure 1-10, each trivalent atom shares its valence atoms with three adjacent atoms in the semiconductor crystal lattice structure. However, the fourth tetravalent atom does not share a covalent bond because of the missing electron. This results in a hole in the crystal lattice. A large number of holes are present in the semiconductor because many trivalent atoms have been added. These holes readily accept electrons from other atoms. However, when a given hole is filled by an electron from another atom, the electron leaves another hole. Therefore, the holes drift from one covalent bond to another in the direction opposite to that of electron movement. Consequently, holes behave like positively charged particles. The acceptor atoms remain fixed within the crystal lattice even though the holes can move freely.

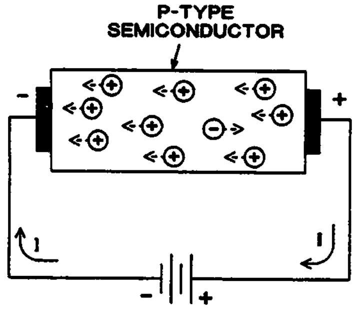


Figure 1-11
Conduction in a
P-type Semiconductor.

If a voltage is applied to a P-type semiconductor as shown in Figure 1-11, the holes provided by the acceptor atoms move from the positive to the negative terminal. These holes move in the same manner previously described. In other words as each electron moves into a hole, another hole is created in its place and since the electrons are attracted to the positive terminal, the holes move in the opposite direction.

In addition to the holes provided by the acceptor atoms, many additional holes are also found in the P-type semiconductor material. These holes are produced as electrons break away from covalent bonds to create electron-hole pairs and are, therefore, produced in the same manner as the holes which appear in a pure semiconductor. These additional holes are also attracted toward the negative terminal while the corresponding electrons that are produced are attracted toward the positive terminal.

Under normal conditions, the number of holes provided by the acceptor atoms will greatly exceed the number of holes and electrons that are produced by the breaking of covalent bonds. The number of holes flowing in the P-type semiconductor will therefore, greatly exceed the free electrons in the material. The holes being in the majority are referred to as the majority carriers and the electrons, which are in the minority, are referred to as the minority carriers.

It is important to note that the semiconductor materials just described are referred to as N-type and P-type semiconductors because the majority carriers within these materials are electrons and holes which respectively have negative and positive charges. The N-type and P-type materials themselves are not charged. In fact, both materials are electrically neutral or uncharged. This is because each atom within each type of material has an equal number of electrons and protons. Also, the movement of holes and electrons does not cause the materials to become charged.

It is also important to understand that N-type and P-type semiconductors have a much higher conductivity than pure semiconductors. Also, the conductivity of these materials can be increased or decreased by simply adding more or less impurities. In other words, the more heavily a semiconductor is doped, the lower its electrical resistance.

## Self-Test Review

39. The conductivity of a semiconductor can be increased by process known as ______________.

40. A pure semiconductor is doped by adding ______________.

41. An impurity material which is made up of atoms that have five valence electrons is called a ______________ material.

42. An impurity material which is made up of atoms that have three valence electrons is called a ______________ material.

43. Donor atoms add ______________ to semiconductor crystals.

44. Acceptor atoms create ______________ in semiconductor crystals.

45. In an N-type semiconductor the majority carriers are ____________.

46. In a P-type semiconductor the majority carriers are ______________.

47. Increased doping causes the electrical resistance of a semiconductor to ____________.

## Answers

39. doping

40. impurities

41. pentavalent

42. trivalent

43. electrons

44. holes

45. electrons

46. holes

47. decrease

# UNIT SUMMARY

Semiconductors are substances that are both poor conductors and poor insulators. The most important semiconductors are the chemical elements that have four electrons in the outermost, or valence, shell of their atoms. The two semiconductors used in the manufacture of solid-state components are the elements silicon and germanium; the use of silicon is more common. It is a characteristic of tetravalent atoms to readily share valence electrons with other atoms. This sharing of valence atoms is called a covalent bond. These covalent bonds cause the atoms to form a crystal lattice.

Pure (intrinsic) semiconductors function as insulators at low temperatures because valence electrons are held tightly in their shells. However, as temperature increases, valence electrons are occasionally able to break their covalent bonds and become free electrons, leaving holes in the crystal lattice. If a voltage is applied under these conditions, electrons can move in one direction while holes move in the opposite direction, allowing the semiconductor to act as a fairly good conductor.

Intrinsic semiconductors have limited use in electronics. However, they can be modified, by doping, to meet specific electrical requirements. Doping is the deliberate addition of an impurity having the desired characteristics to the semiconductor.

Doping a semiconductor with a pentavalent element, such as arsenic, adds a large number of free electrons to the crystal lattice. These electrons freely move from atom to atom in one direction within the crystal lattice when a voltage is applied. Hole movement is also present in the crystal lattice, but the quantity of holes is much smaller than the quantity of electrons. For this reason electrons are the majority carriers of current and holes are the minority carriers in a semiconductor that is doped with pentavalent atoms. Semiconductors that are doped with pentavalent atoms are N-Type semiconductors.

Doping a semiconductor with a trivalent element, such as gallium produces a P-Type semiconductor. The valence electrons in the outer shell of each trivalent atom form covalent bonds with three of the four adjacent tetravalent atoms in the crystal lattice. However, the absence of a fourth valence electron in the trivalent atom causes a hole in the crystal lattice. The large number of trivalent atoms in the crystal lattice causes a large number of holes; each hole is capable of accepting a free electron. However, the number of holes greatly outnumbers the free electrons. Consequently, holes are the majority current carriers and electrons are the minority carriers in P-Type semiconductors.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear after the exam. Do not refer back to the lesson material to find the answers. Let the exam be a fair determination of your knowledge. Answer all questions before you check the answers.

1. A typical atom within a pure semiconductor material will have how many electrons in its valence shell?

   A. 2
   B. 3
   C. 4
   D. 8

2. Components which are made from semiconductor materials are often called:

   A. low-voltage components.
   B. passive components.
   C. high-voltage components.
   D. solid-state components.

3. By using semiconductor materials it is possible to construct electronic components that:

   A. can withstand extremely high voltages.
   B. are small but not highly reliable.
   C. are suitable for use in portable or compact electronic equipment because they are small and consume less power.
   D. are highly reliable, are not sensitive to temperature changes, and consume more power than tubes.

4. The atoms within a pure semiconductor are held together:

   A. by covalent bonds.
   B. by their positive and negative charges.
   C. in a random manner.
   D. like the atoms within a conductor

5. Intrinsic semiconductors can be made to perform as conductors or insulators by:

   A. subjecting them to high voltages.
   B. varying the temperature of the material to extreme limits.
   C. controlling the amount of impurities added.
   D. subjecting them to either a positive or a negative voltage.

6. The main reason for doping a semiconductor is to:

   A. make it physically stronger.
   B. make it insensitive to temperature changes.
   C. decrease its conductivity.
   D. increase its conductivity.

7. A hole behaves like a:

   A. positively charged particle.
   B. negatively charged particle.
   C. donor atom.
   D. impurity atom.

8. Current flow in a semiconductor consists of:

   A. electrons only.
   B. holes only.
   C. electrons and holes.
   D. ions.

9. A pure semiconductor is converted to a N-type semiconductor when it is doped with:

   A. a pentavalent material.
   B. a trivalent material.
   C. either a pentavalent or a trivalent material.
   D. impurity atoms.

10. The number of holes and electrons flowing in a semiconductor material are equal when the material is:

   A. doped with a pentavalent impurity.
   B. doped with a trivalent impurity.
   C. doped with either a pentavalent or a trivalent impurity.
   D. in its intrinsic form.

11. The most popular and widely used semiconductor material is:

    A. germanium.
    B. silicon.
    C. carbon.
    D. arsenic.

12. The type of semiconductor material which utilizes electrons as majority carriers is:

    A. N-type.
    B. P-type.
    C. intrinsic.
    D. trivalent.

13. What sophisticated semiconductor component can replace entire electronic circuits?

    A. diode
    B. transistor
    C. thyristor
    D. integrated circuit

14. In an electronic circuit, a semiconductor device:

    A. replaces discrete components like resistor and capacitors.
    B. controls the flow of current to produce a predetermined effect.
    C. can directly replace vacuum tubes.
    D. compensates for various deficiencies.

15. Which of the following is *not* a benefit of semiconductor devices?

    A. lower cost
    B. less power consumption
    C. higher reliability
    D. higher operating voltages
    E. smaller size, less weight

# EXAMINATION ANSWERS

1. C — Pure semiconductor materials such as germanium and silicon are made up of atoms which have four valence or outermost electrons. These four valence electrons cause the semiconductor materials to have electrical characteristics that are intermediate to those of insulators and conductors.

2. D — Components made from semiconductors are often called solid-state components because they are constructed from a solid material.

3. C — With semiconductor materials it is possible to construct extremely small components which consume very little power and are therefore suitable for use in compact or portable equipment.

4. A — Each atom within the semiconductor material shares its valence electrons with four adjacent atoms to form a covalent bond.

5. B — At extremely high temperatures a semiconductor will function as a conductor, but at extremely low temperatures it will function as an insulator.

6. D — The conductivity of a semiconductor is greatly increased by doping it with either a pentavalent or trivalent material.

7. A — A hole represents a vacant space that was once occupied by an electron, but it has the effect of a positive charge and is able to move around like an electron.

8. C — Both hole flow and electron flow must be considered when analyzing the electrical characteristics of semiconductors.

9. A — The pentavalent material adds extra electrons to the pure semiconductor to produce a negative or N-type semiconductor.

10. D— An intrinsic or pure semiconductor will have an equal number of electrons and holes. These holes and electrons are produced in pairs and the quantity produced is greatly affected by temperature.

11. B— The most popular and widely used semiconductor is silicon because of its superior temperature characteristics.

12 A— The type of semiconductor material that has electrons as the majority carriers is N-type material which can be either silicon or germanium.

13. D— An integrated circuit can replace entire electronic circuits and in many cases entire systems, functions or pieces of equipment.

14. B— The whole purpose of a semiconductor device is to control an electrical signal to produce some desired end result or function.

15. D— This is *not* an advantage or benefit of semiconductor devices.

## *Unit 2*

# SEMICONDUCTOR DIODES

# CONTENTS

Introduction.............................................Page 2-4

Unit Objectives.........................................Page 2-4

Unit Activity Guide....................................Page 2-5

The PN Junction.......................................Page 2-6

- Ions..................................................Page 2-6

- Junction Diodes......................................Page 2-7

  - Depletion Region.................................Page 2-7

  - Barrier Voltage...................................Page 2-8

Diode Biasing..........................................Page 2-10

- Forward Bias.........................................Page 2-10

- Reverse Bias.........................................Page 2-12

Diode Characteristics.................................Page 2-16

- Gemanium Diode.....................................Page 2-17

  - Forward Characteristics.........................Page 2-17

  - Reverse Characteristics.........................Page 2-18

- Silicon Diode........................................Page 2-18

  - Forward Characteristics.........................Page 2-19

  - Reverse Characteristics.........................Page 2-19

- Diode Ratings........................................Page 2-20

- Temperature Considerations.........................Page 2-20

- Diode Symbols.......................................Page 2-21

Experiment 1 .......... Page 2-24

Diode Construction .......... Page 2-30

  Grown Junctions .......... Page 2-30

  Alloyed Junctions .......... Page 2-31

  Diffused Junctions .......... Page 2-31

  Diode Packaging .......... Page 2-32

Experiment 2 .......... Page 2-35

Diode Applications .......... Page 2-38

  Rectificaton .......... Page 2-38

  Other Applications for Diodes .......... Page 2-39

Unit Summary .......... Page 2-41

Unit Examination .......... Page 2-43

# INTRODUCTION

In the previous Unit, you examined the semiconductor materials that are used to construct various types of solid-state components. Now you will learn how these materials are used to construct one of the most important types of solid-state components, the semiconductor diode. Although simple in construction and operation, semiconductor diodes are widely used in many types of electronic equipment. Rectification, waveshaping, circuit protection, and logic operations are only a few of the diversified applications. In addition, there are a variety of diode types each of which is optimized for a particular job. In this Unit, you will learn semiconductor diode fundamentals. In later units, you will learn about the many different diode types and their applications.

Since transistors, integrated circuits, and other solid-state components are constructed in basically the same manner as diodes, an understanding of diode construction and operation is an essential first-step in understanding semiconductor devices.

Examine the Unit Objectives closely and then follow the instructions in the Unit Activity Guide to be sure that you perform all the steps necessary to complete this unit. Keep track of the time you spend on each section and note it in the spaces provided.

# UNIT OBJECTIVES

When you have completed this unit you should be able to:

1. Describe the construction of a semiconductor diode PN junction.
2. Recognize the schematic symbol of a semiconductor diode.
3. Describe the difference between an atom and an ion.
4. Describe the electrical characteristics of a PN junction.
5. Identify the two parts of a diode on the schematic symbol and on an actual diode.
6. Describe the effects of forward and reverse bias characteristics of a junction diode.
7. Interpret a graph of diode voltage-current characteristics.
8. Use an ohmmeter to determine if a diode PN junction is operational or faulty.
9. List at least four applications for semiconductor diodes.
10. Identify at least two different diode packages.

# UNIT ACTIVITY GUIDE

**Completion Time**

☐ Read "The PN Junction." __________

☐ Answer Self Test Review Questions 1-8. __________

☐ Read "Diode Biasing." __________

☐ Answer Self Test Review Questions 9-22. __________

☐ Read "Diode Characteristics." __________

☐ Answer Self Test Review Questions 23-34. __________

☐ Perform Experiment 1 on "Semiconductor Diode Characteristics." __________

☐ Read "Diode Construction. __________

☐ Answer Self Test Review Questions 35-42. __________

☐ Perform Experiment 2 on "Testing Semiconductor Diodes." __________

☐ Read "Diode Applications." __________

☐ Answer Self Test Review Questions 43-46. __________

☐ Read the Unit Summary. __________

☐ Complete the Unit Examination. __________

☐ Check the Examination Answers. __________

# THE PN JUNCTION

As you learned in the previous Unit, a pure or intrinsic semiconductor can be doped with a pentavalent or trivalent material to obtain two basic types of semiconductor materials. These doped semiconductors are referred to as N-type or P-type materials because the majority carriers within them are either electrons or holes respectively. The N-type and P-type designations do not imply that the materials have negative and positive charges. In fact, both types are electrically neutral because the atoms within each semiconductor, including the impurity atoms, contribute an equal number of protons and electrons to the respective materials. Even though some of the electrons and holes are free to move around in a random manner, the semiconductor materials do not acquire an overall electrical charge.

## Ions

Even though N-type and P-type semiconductors are electrically neutral, within each type of semiconductor material, independent electrical charges do exist. For example, each time an atom gives up an electron, it loses a negative charge and becomes electrically unbalanced. In other words, the positive protons in the nucleus of the atom outnumber the electrons revolving around the nucleus and the atom takes on a positive charge. Also when a neutral atom takes on an additional electron, it has more negative electrons than positive protons and therefore assumes a negative charge. These electrically charged atoms are referred to as *ions*. A positively charged atom is therefore referred to as a positive ion and a negatively charged atom is referred to as a negative ion.

The free electrons and holes which drift throughout the semiconductors also possess negative and positive charges respectively. However, since these charges are able to move, they are referred to as *mobile charges*. Within a N-type or P-type semiconductor, an equal number of mobile charges and ionic charges will exist and since these charges are equal and opposite, the semiconductor material remains electrically neutral.

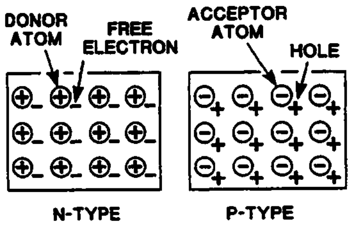


Figure 2-1
Doped semiconductors.

The internal structure of N-type and P-type semiconductors may be illustrated in a simplified manner as shown in Figure 2-1. The N-type semiconductor is doped with a pentavalent impurity and therefore contains many donor atoms which contribute free electrons that can drift through the material. The donor atoms take on positive charges and become positive ions when they release free electrons. These are represented by plus signs surrounded by small circles as shown. The free electrons which accompany the donor atoms are represented by minus signs.

The P-type semiconductor is doped with a trivalent impurity and therefore contains many acceptor atoms. These atoms easily accept or absorb electrons from the semiconductor material and become negatively charged atoms or negative ions. The negatively charged acceptor atoms are represented by minus signs that are surrounded by small circles as shown. The holes created by these atoms act like positively charged particles and are represented by plus signs as indicated. By representing the doped semiconductors in this manner it becomes relatively easy to analyze the action which takes place when they are combined to form diodes and other types of solid-state components.

# Junction Diodes

Now let's consider the action which takes place when doped semiconductors are combined to form a diode. Basically a diode is created by joining N-type and P-type semiconductors as shown in Figure 2-2. When these oppositely doped materials come in contact with each other, a junction is formed where they meet. Such diodes are referred to as *junction diodes* and can be made of either silicon or germanium. When the junction is formed, a unique action takes place.

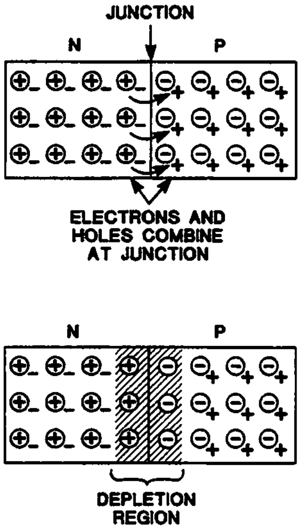


Figure 2-2
Characteristics of a PN Junction

## Depletion Region

The mobile charges, free electrons and holes, in the vicinity of the junction are strongly attracted to each other and therefore drift toward the junction. The accumulated charges at each side of the junction serve to increase the attraction even more. Eventually some of the free electrons move across the junction and fill some of the holes in the P-type material. As the free electrons cross the junction, the N-type material becomes depleted of electrons in the vicinity of the junction. At the same time, the holes within the P-type material become filled and no longer exist. This means that the P-type material also becomes depleted of holes near the junction. This region near the junction where the electrons and holes become depleted is referred to as the *depletion region.* The depletion region extends for only a short distance on each side of the junction as shown in Figure 2-2.

It is important to remember that free electrons and holes are actually the majority carriers for the N-type and P-type materials. Therefore, no majority carriers exist within the depletion region. Also, it is important to note that the N-type and P-type materials are no longer neutral or uncharged. In other words the N-type material has lost free electrons thus causing the positive donor atoms (ions) to outnumber the free electrons. The N-type material therefore takes on a positive charge near the junction. The P-type material has lost holes which means that the negatively charged acceptor atoms (ions) will outnumber the holes. The P-type material therefore, takes on a negative charge near the junction. This means that opposite charges now exist on each side of the junction.

The depletion region does not continue to become larger and larger until the N-type and P-type materials are completely depleted of majority carriers. Instead, the action of the electrons and holes combining at the junction tapers off very quickly. Therefore, the depletion area remains relatively small. The size of the depletion region is limited by the opposite charges which build up on each side of the junction. The negative charge which accumulates in the P-type material eventually becomes great enough to repel the free electrons and prevent them from crossing the junction. The positive charge which accumulates in the N-type material also helps to stop the flow of free electrons by attracting and holding them back so they cannot move across the junction.

## Barrier Voltage

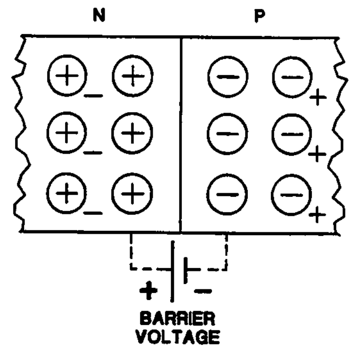


Figure 2-3

Barrier Voltage in a PN Junction.

The opposite charges that build up on each side of the junction create a difference in potential or voltage. This difference in potential effectively limits the size of the depletion region by preventing the further combination of electrons and holes. It is referred to as the *potential barrier* or the *barrier voltage*. Although this barrier voltage exists across the PN junction, its effect can be represented by an external battery as shown in Figure 2-3.

The barrier voltage produced within a PN junction will usually be on the order of several tenths of a volt. This voltage will always be higher for silicon PN junctions than for germanium PN junctions. For example, a PN junction made from doped germanium will have a typical barrier voltage of 0.3 volts while a PN junction made from doped silicon will have a typical barrier voltage of 0.7 volts. Although the barrier voltage exists inside of the junction diode and therefore cannot be measured directly, its presence becomes apparent when an external voltage is applied to the diode. Later in this unit you will see why this is true.

# Self-Test Review

1. Electrically charged atoms are referred to as ______________.

2. The free electrons and holes which drift throughout a semiconductor are referred to as ______________ ______________.

3. A diode is formed by joining ______________ and ______________ semiconductor materials.

4. When a PN junction is formed, a ______________ region is created in the area near the junction.

5. The opposite charges which build up on each side of a PN junction create a difference of potential which is known as the

   a. mobile charge
   b. bias voltage
   c. barrier voltage

6. The barrier voltage with a diode will usually be in which of the following ranges?

   a. 10 to 100 mV
   b. 100 mV to 1 V
   c. 1 V to 10 V
   d. 10 V to 100 V

7. A voltmeter may be used to directly measure the barrier voltage within a diode.

   a. True
   b. False

8. An understanding of diode operation can only be obtained by examining the action which takes place at the ______________ of the diode.

## Answers

1. ions.
2. mobile charges
3. N-type and P-type
4. depletion
5. (c) barrier voltage
6. (b) 100 mV to 1 V range
7. (b) False — the barrier voltage exists inside of the diode and cannot be directly measured.
8. junction

# DIODE BIASING

Whenever diodes are used in electronic circuits, they are subjected to various voltages and currents. The polarities and amplitudes of the voltages and currents must be such that proper diode action takes place. We generally refer to the voltages applied to a semiconductor diode as bias voltages. Let's see how these voltages affect and control the diode's operation.

## Forward Bias

In the previous discussion, you saw that free electrons and holes, the majority carriers, combine at the junction to produce a depletion region. The depletion region represents an area that is void of majority carriers but at the same time contains a number of positively charged donor atoms (ions) and negatively charged acceptor atoms (ions). These positive and negative charges are separated at the junction and effectively create a barrier voltage which opposes any further combining of majority carriers. It is important to remember that this action takes place in a PN junction diode that is not subjected to any external voltage.

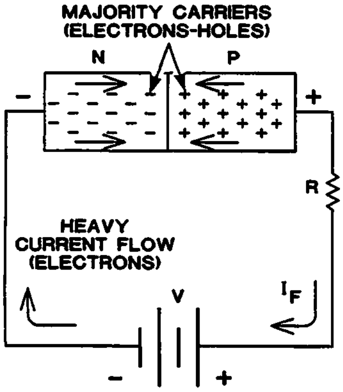


Figure 2-4
Forward-biased
PN Junction Diode

When a PN junction diode is subjected to a sufficiently high external voltage V as shown in Figure 2-4, the device will function in a somewhat different manner. Notice that the negative and positive terminals of a battery are connected to the N and P sections of the diode respectively. An external resistor R is used to limit the current level to a safe value. Under these conditions the free electrons in the N section are repelled by the negative battery terminal and forced toward the junction where they will neutralize the positively charged donor atoms (positive ions) in the depletion region. During this same period of time, the free electrons that had initially accumulated to create a negative charge on the P side of the junction are attracted toward the positive battery terminal. Therefore the negative charge on the P side of the junction is also neutralized. This means that the positive and negative charges which form the internal barrier voltage are effectively neutralized and no barrier voltage will be present to stop the combining of majority carriers at the junction. The PN junction diode is therefore able to support a continuous flow of current at this time. This action will occur only if the battery voltage V is greater than the barrier potential.

Since the diode is now subjected to an external voltage, a constant supply of electrons flow into the N section of the diode. These electrons drift through the N-type material toward the junction. The movement of these electrons through the N section is sustained by the free electrons (majority carriers) that exist within this material. At the same time, the holes (majority carriers) in the P section also drift toward the junction. The electrons and holes combine at the junction and effectively disappear as they neutralize each other. However, as these electrons and holes combine and are effectively eliminated as charge carriers, new electrons and holes appear at the outer edges of the N and P sections. The majority carriers therefore continue to move toward the junction as long as the external voltage is applied. This action is shown in Figure 2-4.

It is convenient to analyze the action which takes place in the P section of the diode by considering the movement of holes instead of electrons. However, it is important to realize that electrons do actually flow through the P-type material. The electrons are attracted by the positive terminal of the battery and as the electrons leave the P section and enter the battery, holes are created at the outer edge of the P section. These holes drift toward the junction where they combine with electrons and effectively disappear.

The important point to note at this time is that electrons do flow through the entire PN junction diode when it is subjected to a sufficiently high external voltage. At this time the diode is said to be conducting current in the *forward* direction. Also the diode is considered to be *forward-biased* by the external voltage. The current which flows through the forward-biased diode shown in Figure 2-4 is limited by the resistance of the P-type and N-type semiconductor materials as well as the external resistance R. Normally the diode resistance is quite low. Connecting a forward bias voltage directly to the diode will result in a current large enough to generate sufficient heat to destroy the diode. For this reason, forward-biased diodes are usually connected in series with a resistor or some other device which will limit the current to a safe level.

A forward-biased diode will conduct current as long as the external bias voltage is sufficiently high and the polarity is correct. For example, if the diode is constructed from germanium a forward bias of approximately 0.3 volts will be required before the diode can begin to conduct. Silicon diodes require a forward bias of approximately 0.7 volts in order to begin conducting. The external voltage applied to the diode must be large enough to neutralize the depletion area and therefore neutralize the barrier voltage that exists across the PN junction of the diode. Once this internal voltage is overcome, the diode will conduct in the forward direction.

The polarity of the external dc bias voltage must also be correct with respect to the P and N sections of the diode. The negative terminal of the bias source must be connected to the N section and the positive terminal should be connected to the P section to achieve the forward-biased condition.

Once the diode is conducting, a voltage will be dropped across the device. This occurs because the diode's semiconductor material has a low but finite resistance value and the current flowing through it must produce a corresponding voltage drop. As it turns out, this forward bias voltage drop is approximately equal to the barrier potential. This is 0.3 volts for a germanium diode and 0.7 volts for a silicon diode.

The amount of forward bias current $I_F$ is a function of the applied dc bias V, the forward voltage drop $V_F$, and the external resistance R. The relationship simply involves Ohm's law as indicated below

$$I_F = \frac{V - V_F}{R}$$

For example, the forward current in a silicon diode with a bias voltage of 10 volts and an external resistor of 100 ohms is

$$I_F = \frac{10 - .7}{100} = \frac{9.3}{100} = .093 \text{ amps or } 93 \text{ ma}$$

## Reverse Bias

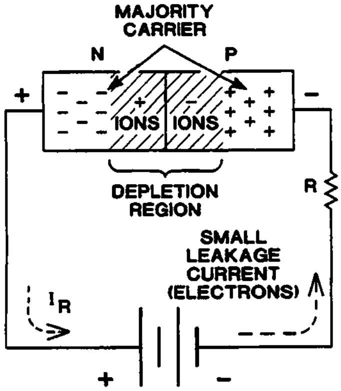


Figure 2-5
Reversed-biased PN Junction diode.

A forward-biased diode is able to conduct current in the forward direction because the external bias voltage forces the majority carriers together so that they can combine at the junction of the diode and create a continuous flow of current. In order to achieve this condition the negative terminal of the battery is connected to the N section of the diode and the positive terminal is connected to the P section. However, if the battery connections are reversed as shown in Figure 2-5, the diode will operate in a different manner. The negative terminal of the battery is now connected to the P section of the diode while the positive terminal is connected to the N section. The diode is now considered to be *reverse-biased.* Under these conditions the free electrons in the N section will be attracted toward the positive battery terminal thus leaving a relatively large number of positively charged donor atoms (positive ions) in the vicinity of the junction as shown in Figure 2-5. In fact, the number of positive ions in the N section at times will even outnumber the positive ions that exist in an unbiased diode. This effectively increases the width of the depletion region on the N side of the junction thus causing the positive charge on this side of the junction to increase. At the same time, a number of electrons leave the negative terminal of the battery and enter the P section of the diode. These electrons fill the holes near the junction thus causing the holes to move toward the negative terminal. A large number of negatively charged acceptor atoms (negative ions) are therefore created near the junction. This effectively increases the width of the depletion region on the P side of the junction.

The overall depletion region of the diode shown in figure 2-5 is wider than the depletion region in the unbiased diode shown in Figure 2-2. This means that the opposite charges on each side of the junction are also larger and therefore create a higher barrier voltage across the junction. These opposite charges will build up until the internal barrier voltage is equal and opposite to the external battery voltage. Under these conditions the holes and electrons (majority carriers) cannot support current flow and the diode effectively stops conducting.

Actually, an extremely small current will flow through the diode shown in Figure 2-5. This small current is sometimes referred to as a *leakage current* or a *reverse current* and is designated as $I_R$. It exists because of the minority carriers which are contained within the N and P sections of the diode. Remember that the minority carriers are holes in the N-type material and electrons in the P-type material. When the diode is reverse-biased as shown in Figure 2-5 the minority carriers are forced toward the junction where they combine and thus support an extremely small current. This action closely resembles the action which takes place in the forward-biased diode shown in Figure 2-4 but it is on a much smaller scale.

The number of minority carriers in the N and P materials is extremely small at room temperature. However, as temperature increases, a greater number of electron-hole pairs are generated within the two materials. This causes an increase in minority carriers and a corresponding increase in leakage current.

All PN junction diodes produce some leakage current when they are reverse-biased but this current is very small. It is only microamperes in germanium diodes and nanoamperes in silicon diodes. These currents are usually many orders of magnitude less than the usual forward current. It is important to remember that germanium diodes normally produce a high leakage current than silicon diodes. Germanium devices are also more temperature sensitive than silicon devices. This disadvantage of germanium diodes is often offset by the desirable lower barrier potential and forward voltage drop.

We can sum up the operation of a PN junction diode in this manner. The diode is a unidirectional electrical device since it conducts current in only one direction. When it is forward biased, current flows through it freely since it acts as a very low resistance. When the diode is reversed biased, current does not flow through it. It simply acts as an open circuit or extremely high resistance. Only a small, temperature sensitive leakage current flows in the reverse biased condition. The diode is effectively a polarity sensitive electrical switch. When forward biased the diode switch is closed. When reverse biased it is open.

## Self Test Review

9. The depletion region is neutralized when a diode is ________________.

10. A diode is forward-biased when the negative terminal of the external battery is connected to the ________________ section of the diode.

11. A forward-biased diode acts like a switch that is ________________.

12. The current flowing in a forward-biased diode is sustained by the ________________ carriers in the N and P materials.

13. The amount of forward-bias voltage applied to a diode before it can conduct must be greater than the ____________________________.

14. A reverse-biased diode acts like a switch that is ________________.

15. A diode is reverse-biased when the negative terminal of the external battery is connected to the ______________ section of the diode.

16. When a diode is reversed-biased, its depletion area is ______________ than when it is unbiased.

17. The current that flows through a reversed-biased diode is sustained by the ________________ carriers in the N and P materials.

18. The barrier voltage produced within a reverse-biased diode is equal and opposite to the external voltage applied.

    a. True
    b. False

19. The small reverse current that flows through a reverse-biased diode is sometimes called a ____________ current.

20. As temperature increases, the reverse current in a diode __________.

21. An external source of 5 volts dc is applied to germanium diode in series with a 1k ohm resistor. The positive source lead is connected to the P section of the diode. How much current flows, if any? ____________

22. The approximate voltage drop across a silicon PN junction is ___________ volt.

## Answers

9. forward-biased
10. N
11. closed
12. majority
13. barrier potential
14. open
15. P
16. larger or wider
17. minority
18. (a) true
19. leakage
20. increases
21. The value of $I_F$ can be determined in the following manner.

$$I_F = \frac{V - V_F}{R} = \frac{5 - .3}{1000} = \frac{4.7}{1000}$$

$$= .0047 \text{ amps or } 4.7 \text{ ma}$$

22. 0.7 volt

# DIODE CHARACTERISTICS

Now that you have seen how a basic PN junction diode operates, you are ready to examine some of the important electrical characteristics of these devices. Since the characteristics of diodes vary considerably when they are subjected to various voltages and temperatures, it is usually best to plot the desired characteristics in a graphical manner. This makes it possible to analyze the operation of the device at a variety of points where the voltages, currents, or temperatures involved have specific and related values.

The graph in Figure 2-6 shows the amount of forward current and reverse current that flows through a typical PN junction diode when the device is first forward-biased and then reversed-biased. The diode's forward and reverse bias voltages, $V_F$ and $V_R$, are plotted to the right and to the left respectively on the horizontal axis of the graph. The diode's forward and reverse currents, $I_F$ and $I_R$, are plotted above and below the horizontal axis respectively to form the vertical axis of the graph. The point where the horizontal axis crosses the vertical axis is often referred to as the origin of the graph. This origin serves as a zero reference point for the four quantities involved. A graph like the one shown in Figure 2-6 is created by actually subjecting a diode to various forward and reverse voltages while measuring the current through the diode. However, certain precautions must be taken to insure that the diode is not damaged by excessive current or voltage.

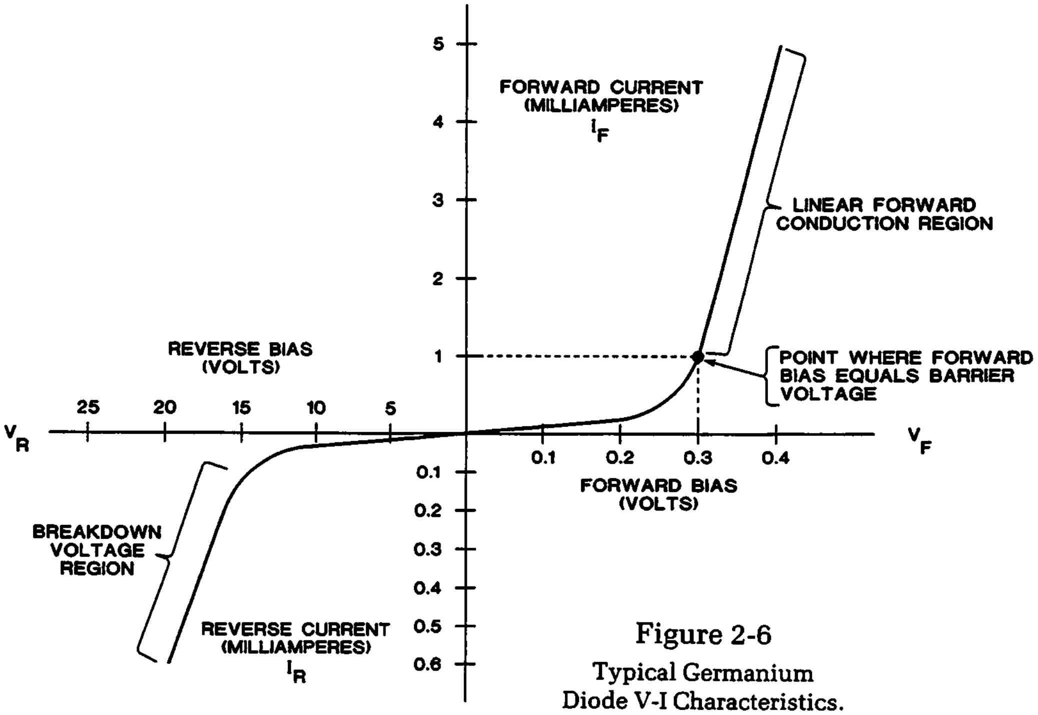


Figure 2-6

Typical Germanium Diode V-I Characteristics.

When a large number of corresponding voltage and current values are plotted, a continuous curve is obtained as shown. For this reason, a graph like the one shown in Figure 2-6 is often referred to as a voltage-current or V-I characteristic curve. If you examine Figure 2-6 closely you will find that the forward currents and voltages and the reverse currents and voltages are plotted to different scales. This is because the forward characteristics involve low voltages and high currents while the reverse characteristics involve relatively high voltages and low currents.

## Germanium Diode

The V-I characteristic curve in Figure 2-6 is for a germanium diode. Let's consider its operation in detail.

### Forward characteristics

Figure 2-6 shows that the forward current through a germanium diode is extremely small and almost insignificant until the forward bias voltage across the diode increases beyond a value of 0.2 volts. Then the forward current increases as the forward bias voltage is increased still further. The increase in forward current really starts to occur as the external bias voltage overcomes the diode's internal barrier voltage. Once the bias voltage exceeds the barrier voltage (0.3 volts), the forward current increases very rapidly and at a linear rate because the diode then acts as a low resistance. If this forward current continued to rise, the diode would eventually be damaged by an excessive flow of current. Throughout the linear portion of the curve, the voltage across the diode is only several tenths of a volt as shown. While the forward voltage drop is not constant, it changes very little over a wide current range. A tremendous change in forward current occurs while the voltage across the diode changes only a small amount.

The point at which the bias voltage equals the barrier voltage is indicated in Figure 2-6. Notice that this point occurs when the bias voltage is equal to 0.3 volts. Also notice that this point occurs when the bias voltage is equal to 0.3 volts. Also notice that the diode's forward current is equal to 1 milliampere at this time and that this current can increase above 5 milliamperes while corresponding voltage across the diode remains below 0.4 volts. Figure 2-6 therefore shows that the diode's internal barrier voltage is approximately 0.3 volts. However, it is important to realize that this voltage will vary from one germanium diode to the next.

### Reverse Characteristics

The V-I curve in Figure 2-6 also shows that when the diode is reverse-biased, the reverse current that flows is extremely small. Notice that the reverse current increases slightly as the reverse voltage increases but remains less than 0.1 milliamperes (100 microamperes) until the reverse voltage approaches a value of 20 volts. Then the reverse current suddenly increases to a much higher value. This sudden increase in reverse current results because the reverse bias voltage becomes strong enough to tear many valence electrons from their parent atoms and therefore increase the number of electron-hole pairs in the N and P materials. This causes an increase in minority carriers which in turn support a higher reverse current. In other words the junction simply breaks down when the reverse bias voltage approaches a value of 20 volts.

The voltage at which the sudden change occurs is commonly referred to as the *breakdown voltage*. This breakdown voltage will vary from one diode to the next since it is determined by the exact manner in which the diode is constructed. In certain cases, ordinary germanium diodes can be damaged when breakdown occurs; however, there are special diodes which are designed to operate in this region. These special devices, known as zener diodes, will be described in detail in the next Unit. When breakdown occurs, the diode no longer offers a high resistance to the flow of reverse current and therefore cannot effectively block current in the reverse direction. For these reasons, operation in the breakdown region is avoided when an ordinary PN junction diode is being used.

## Silicon Diode

While a silicon diode operates the same as the germanium diode, there are some important differences in their characteristic curves. Let's look at these differences in detail.

## Forward Characteristics

The V-I curve in Figure 2-7 shows the characteristics of a typical silicon diode. Notice that the forward characteristics of this diode are basically similar to those of the germanium diode previously described; however, there is an important exception. The internal barrier voltage of the silicon diode is not overcome until the forward bias voltage is equal to approximately 0.7 volts as shown. Beyond this point the forward current increases rapidly and at a linear rate. The corresponding forward voltage across the diode increases only slightly. The exact amount of forward voltage required to overcome the barrier voltage will vary from one silicon diode to the next but will usually be close to the 0.7 volts indicated in Figure 2-7.

## Reverse Characteristics

The reverse characteristics of the silicon diode are also similar to those of the typical germanium diode previously described. However, the silicon diode has a much lower reverse current than the germanium type as indicated in Figure 2-7. Notice that the reverse current remains well below 0.1 milliamperes (100 microamperes) until the breakdown voltage of the device is reached. Then, as with the germanium unit, a relatively high reverse current is allowed to flow. A breakdown voltage of 45 volts is indicated in Figure 2-7, however, this voltage will vary from one silicon diode to the next. Also, the reverse currents in many silicon diodes may be in the extremely low nanoampere range and therefore insignificant for most practical applications.

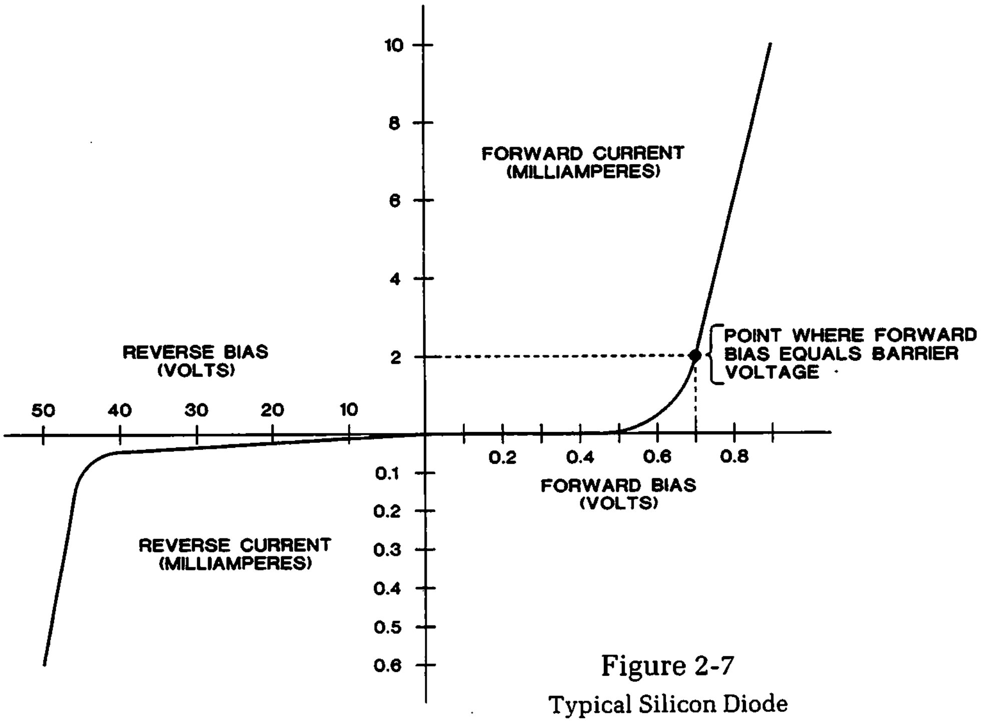


Figure 2-7
Typical Silicon Diode
V-I Characteristics.

# Diode Ratings

When the important characteristics of silicon and germanium diodes are compared it becomes apparent that either type can be damaged by excessive forward current. For this reason, manufacturers of these diodes usually specify the maximum forward current ($I_F$ max) that each type can safely handle. Also, both types can be damaged by excessive reverse voltages which cause the diode to breakdown and conduct a relatively large reverse current. To insure that the various diodes are not subjected to dangerously high voltages, manufacturers of these devices usually specify the maximum safe reverse voltage that can be applied to each particular device. This maximum reverse voltage is commonly referred to as the *peak inverse voltage*, which is usually abbreviated PIV.

# Temperature Considerations

In some critical applications it is also necessary to consider the effect that temperature has on diode operation. In general, the diode characteristic that is most adversely affected by changes in temperature is the diode's reverse current. This current is caused by the minority carriers that are present in the N and P sections of the diode as explained previously. At extremely low temperatures the reverse current through a typical diode will be practically zero. But at room temperature this current will be somewhat higher although still quite small. At extremely high temperatures an even higher reverse current will flow which in some cases might interfere with normal diode operation. These changes in reverse current as a result of temperature changes are shown in Figure 2-8. Notice that the breakdown voltage also tends to increase as temperature increases, however, this change is not great.

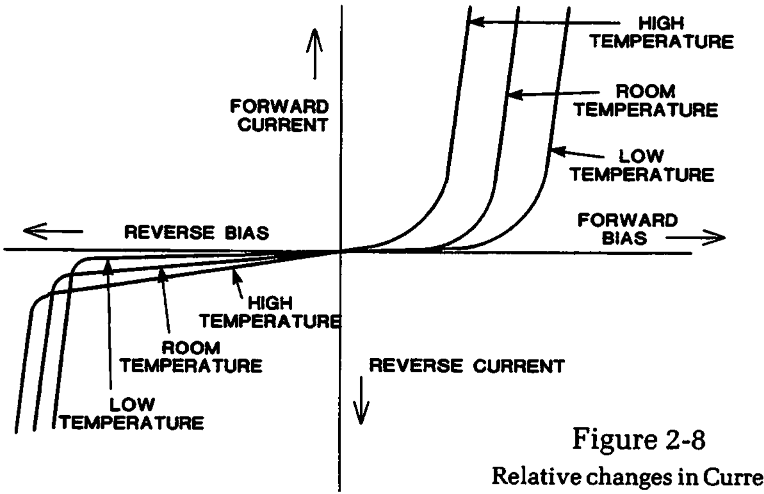


Figure 2-8
Relative changes in Current due to changes in Temperature.

These same relative changes occur in both types of diodes, even though the reverse currents are generally higher in the germanium types. For both germanium and silicon diodes, the reverse or leakage current doubles for approximately every 10° Centigrade rise in temperature.

The forward voltage drop across a conducting diode is also affected by temperature changes. This is illustrated in Figure 2-8. The forward voltage drop is inversely proportional to temperature. As the temperature rises, the voltage drop decreases. This effect is the same in both germanium and silicon devices.

## Diode Symbols

When diodes are shown in a circuit drawing or schematic, it is convenient to represent them with an appropriate symbol. The symbol most commonly used to represent the diode is shown in Figure 2-9 along with a typical PN junction diode. Notice that the P section of the diode is represented by a triangle (also called an arrow) and the N section is represented by a bar (also called a rectangle). The arrows that are placed beside the diode and its symbol indicate the direction of forward current ($I_F$) or electron flow. As you discovered earlier, the forward current must flow from the N section to the P section of the diode. This means that the forward current through the symbol must flow from the bar or rectangle to the triangle or arrow. In other words forward current flow is always against the arrow in the diode symbol. Also notice that the N and P sections of the diode and the corresponding portions of the diode symbol have been identified as the *cathode* and the *anode* respectively. These two terms were once widely used to identify the two principle elements within a vacuum tube diode. However, they are now commonly used to describe the two sections of a junction diode. The cathode (N-type) is simply the section of the diode that supplies the electrons and the anode (P-type) is the section that collects the electrons.

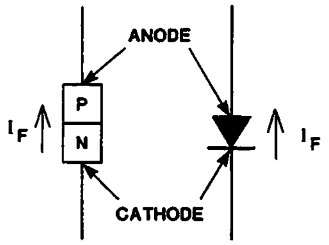


Figure 2-9

A typical Junction Diode and its Symbol.

Figure 2-10 shows how forward-biased and reversed-biased diodes are represented in schematic form. Notice that when the negative and positive terminals of the battery are connected to the cathode and anode of the diode respectively, the diode is forward-biased and will conduct a relatively high forward current. ($I_F$). The resistor is added in series with the diode as shown to limit this forward current to a safe value. Also notice that when the battery terminals are reversed, the diode is reverse-biased and only a very low reverse current ($I_R$) will flow through the device.

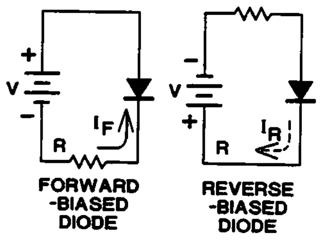


Figure 2-10

Forward and Reverse-biased diode circuits.

## Self Test Review

23. The V-I characteristics curves of silicon and germanium diodes are similar.

    a. True
    b. False

24. A germanium diode will act as a low resistance in the the forward direction when its forward voltage is grater than its internal ____________ voltage.

25. A sudden increase in reverse current results when the ____________ voltage of a diode is reached.

26. Reverse current is usually higher in ____________ diodes.

27. The forward current in a diode is directly proportional (linear) to forward voltage when the forward voltage exceeds the diode's barrier voltage.

    a. True
    b. False

28. The abbreviation PIV stands for ____________ ____________ ____________.

29. The forward voltage drop across a silicon diode will normally be higher than the forward voltage across a germanium diode when both diodes are conducting a relatively high forward current.

    a. True
    b. False

30 Germanium and silicon diodes cannot be damaged by excessively high voltages since no current flows.

    a. True
    b. False

31. As the temperature of a diode decreases, the forward voltage drop

    a. increases
    b. decreases
    c. remains constant

32. If a reverse-biased diode has a leakage current of 10 microamperes at a temperature of 25° centigrade, what will its leakage current be if the temperature increases to 45° centigrade?

33. A diode is forward-biased when the negative terminal of a battery is connected to the cathode of the diode and the positive terminal is connected to the anode of the diode.

    a. True
    b. False

34. The P section of a diode is represented by the ____________ in the diode symbol.

## Answers

23. (a) True
24. barrier
25. breakdown
26. germanium
27. (a) True
28. peak inverse voltage
29. (a) True
30. (b) False
31. (a) increases
32. 40 microamperes. The leakage current doubles for each 10° C of temperature increase.
33. (a) True
34. triangle or arrow

# EXPERIMENT 1

# Semiconductor Diode Characteristics

*OBJECTIVE:* *To investigate the characteristics of typical germanium and silicon semiconductor diodes.*

## Introduction

Semiconductor diodes conduct current well in one direction but very poorly in the other. Their forward resistance is low while their reverse resistance is high. This unilateral characteristic is typical of nearly all semiconductor diodes. However, the specific characteristics of such diodes vary widely. In this experiment, you will investigate the characteristics, in detail, by plotting V-I curves for typical germanium and silicon diodes. Read the entire procedure through before performing this experiment.

## Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — silicon diode (56-28)
1 — germanium diode (56-602)
1 — 1 kilohm resistor (brown, black, red, gold)
1 — 100 kilohm resistor (brown, black, yellow, gold)

## Procedure

1. Plug in your Trainer, but be sure that the unit is turned off at this time.

2. Construct the circuit shown in Figure 2-11, using the 1 kilohm potentiometer on your Trainer, a 1 kilohm resistor, and the germanium diode (56-602). Connect the circuit as shown in Figure 2-12. Remember that the GND (ground) terminal on your Trainer is common to both the positive (+) and negative (–) power supplies. In this experiment, the POS and GND terminals are used just like the positive and negative leads of a battery.

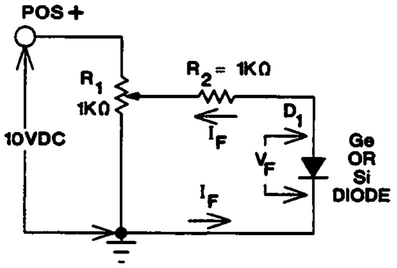


Figure 2-11
Experimental Circuit.

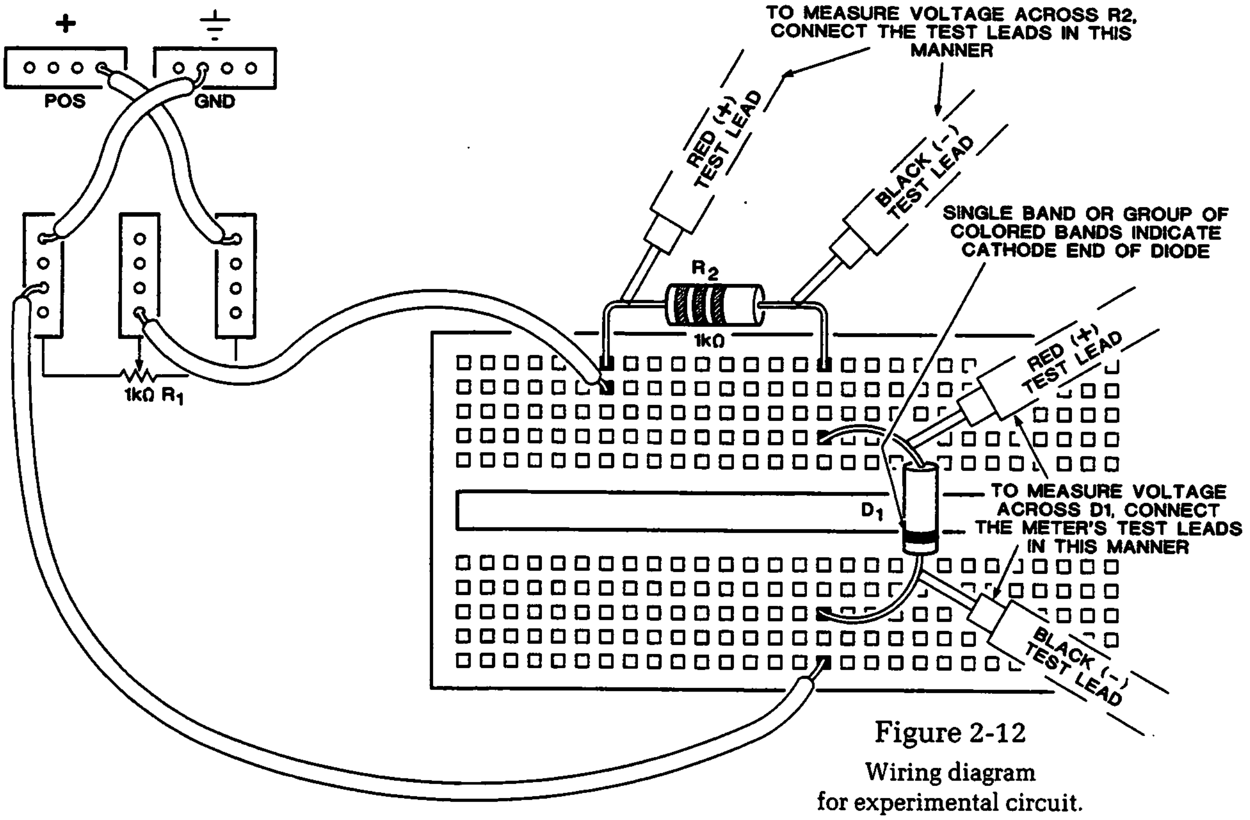

+
POS
GND
1kΩ $R_1$
TO MEASURE VOLTAGE ACROSS R2, CONNECT THE TEST LEADS IN THIS MANNER
RED (+) TEST LEAD
BLACK (–) TEST LEAD
$R_2$
1kΩ
SINGLE BAND OR GROUP OF COLORED BANDS INDICATE CATHODE END OF DIODE
RED (+) TEST LEAD
$D_1$
TO MEASURE VOLTAGE ACROSS D1, CONNECT THE METER'S TEST LEADS IN THIS MANNER
BLACK (–) TEST LEAD

Figure 2-12
Wiring diagram
for experimental circuit.

3. Turn on your Trainer. Turn the positive (+) voltage control clockwise so that 10 volts from the power supply will be applied across the 1 kilohm potentiometer. Measure this voltage with your voltmeter by connecting it across the two end terminals of the 1 kilohm potentiometer. You will use the 1 kilohm potentiometer on the Trainer (designated as $R_1$) to control the voltage applied to the diode ($D_1$) and resistor ($R_2$). This variable voltage will forward bias diode $D_1$. Resistor $R_2$ limits the forward current ($I_F$) flowing through $D_1$ to a safe value.

4. Adjust the diode's forward current to a value of 0.1 milliamperes. To do this, connect your voltmeter across $R_2$ as indicated in Figure 2-12. Then adjust $R_1$ until the meter indicates that the voltage across $R_2$ is equal to 0.1 volts. According to Ohm's law, the voltage across $R_2$ (0.1 volts) divided by the resistance of $R_2$ (1000 ohms) is equal to the current through $R_2$. This current must, therefore, be equal to 0.1 divided by 1000 or 0.0001 amperes, which may also be expressed as 0.1 milliamperes. Since $R_2$ and $D_1$ are in series, the forward current ($I_F$) through $D_1$ must also be equal to 0.1 milliamperes at this time.

5. Without disrupting the setting of $R_1$ use your voltmeter to measure the voltage across $D_1$. Connect the meter's test leads as indicated in Figure 2-12 and note the indicated voltage. Record this voltage in the table in Figure 2-13 immediately below the forward current ($I_F$) value of 0.1 milliamperes. Notice that appropriate spaces are provided for germanium (Ge) and silicon (Si) diodes. Your recorded value represents the forward bias voltage ($V_F$) across the diode when $I_F$ equals 0.1 milliamperes.

| $I_F$(milliamperes) | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $V_F$(volts) - Ge | | | | | | | | | | | |
| $V_F$(volts) - Si | | | | | | | | | | | |

**Figure 2-13**
Chart for recording data
obtained in Steps 5, 6, and 7.

6. Now complete the table in Figure 2-13 by adjusting $I_F$ to the remaining values indicated and record the corresponding values of $V_F$. Use the same basic procedures described in Steps 4 and 5 above, keeping in mind that different values of $I_F$ and $V_F$ are involved. You will save considerable time in completing the table if you simply consider the fact that $I_F$ varies proportionally with the voltage across $R_2$. Therefore, if $I_F$ equals 0.5 milliamperes when the voltage across $R_2$ equals 0.5 volts, $I_F$ values of 1.0 and 2.0 milliamperes will correspond to voltages of 1.0 and 2.0 volts across $R_2$. This makes it possible to convert the voltage across $R_2$ directly into milliamperes. As you complete the table in Figure 2-13, measure each corresponding value of $V_F$ as accurately as possible.

7. Now plot the corresponding values of $I_F$ and $V_F$, in Figure 2-13, on the graph shown in Figure 2-14. Then connect the various points plotted on the graph to form a continuous curve. This curve shows the continuous relationship between the diode's forward current and forward bias voltage. Label this curve Ge, for germanium.

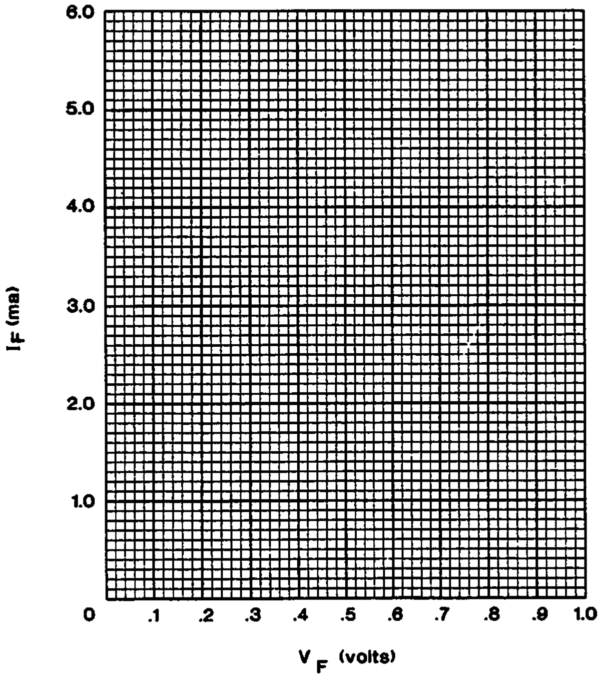


**Figure 2-14**
Graph for plotting $I_F$ and $V_F$ values.

8. Turn off your Trainer and reverse the connections of $D_1$ so that it is installed in the opposite direction. Also, replace $R_2$ with a 100 kilohm resistor. Your circuit should still appear as shown in Figure 2-12, but with the cathode end of $D_1$ now connected to $R_2$, instead to the negative side of the power supply. $D_1$ is now connected so that it will be reverse-biased when voltage is applied to it through $R_2$.

9. Turn on your Trainer, adjust $R_1$ for maximum resistance, and turn the voltage control of the positive (+) power supply to the full clockwise position. You will now use your voltmeter to measure the reverse bias voltage ($V_R$) across $D_1$. Connect the meter's test leads across $D_1$ as shown in Figure 2-12 (since the diode is now reversed, the black or negative lead will now touch the diode's anode and the red or positive lead will touch the diode's cathode). Turn $R_1$ slowly clockwise until the meter indicates that the diode's reverse bias voltage is equal to 12 volts.

10. Without disturbing the setting of $R_1$ use your voltmeter to measure the voltage across $R_2$. Connect the meter's test leads as shown in Figure 2-12 and note the indicated voltage. Use this voltage indication to determine the amount of current flowing through $R_2$.

$$I = \frac{V_{R2}}{R_2} = \frac{V_{R2}}{100{,}000\ \Omega}$$

Since the $R_2$ and $D_1$ are in series and $D_1$ is reverse biased, your calculated current represents the reverse current ($I_R$) flowing through $D_1$. Record this reverse current ($I_R$) value (in microamperes) in the appropriate space in the table in Figure 2-15. Your recorded value represents the amount of reverse current flowing through $D_1$ when $V_R$ equals 12 volts. This reverse current should be very low and for some diodes it may be too low to measure. If you cannot detect any current simply indicate that it is zero. Otherwise try to determine its approximate value.

| $V_R$ - 12 volts | |
|---|---|
| $I_R$ (microamperes) - Ge | |
| $I_R$ (microamperes) - Si | |

**Figure 2-15**
Chart for recording data
obtained in Steps 9 and 10.

11. Turn off your Trainer and remove the germanium diode ($D_1$) from the circuit. Install the silicon diode in place of the germanium diode and replace the 100 kilohm resistor with the 1 kilohm resistor you started with. This will return the circuit to the condition shown in Figure 2-11. Check the wiring diagram in Figure 2-12 to be sure that your circuit is connected properly.

12. Now repeat Steps 3 through 10 using the silicon diode (56-28) instead of the germanium diode. Be sure that you fill in the proper spaces in the tables shown in Figure 2-13 and Figure 2-15. Also, be sure to identify the silicon diode characteristic curve that you plot in Figure 2-14. After you complete Step 10 for the silicon diode, turn off the Trainer, and read the following discussion.

## Discussion

In this experiment, you subjected first a germanium diode and then a silicon diode to various forward bias voltages and adjusted the diode current in steps over the 0.1 mA to 6 mA range. You then determined the corresponding forward voltage across each device. Then, you used these measured values to plot a characteristic curve for each diode. You should have found that the characteristic curves that you plotted for the germanium and silicon diodes compare favorably with the V-I curves shown earlier in this unit (Figures 2-6 and 2-7). Your curves should show that the forward bias voltage across each diode remains relatively constant when the forward current ($I_F$) through each diode reaches a pint where it begins to rise linearly at a rapid rate. This pint is called **the knee** of the curve. At this point, the voltage across the germanium diode was probably in the 0.3 to 0.4 volt range, while the voltage across the silicon diode probably remained in the 0.6 to 0.7 volt range. Below the knee of the V-I curve for each diode, the current and voltage variations are non-linear. However, above the knee of the curve, the current is directly proportional to the forward voltage.

In studying the curves you plotted in Figure 2-14, it should be apparent that the voltage across the germanium diode is less than the voltage across the silicon diode. Also, it should be apparent that the forward voltage is not a fixed value; it does vary in proportion to the forward current. However, the amount of change of the forward voltage over a wide range of current values is very small. For that reason, we can generally assume the forward voltage to be basically a constant for most applications.

Next, you subjected the germanium and silicon diodes to reverse voltages and measured the reverse or leakage currents through these devices. You should have found that the reverse current through each diode was extremely low and for all practical purposes equal to zero. You probably found the leakage current in the germanium diode to be less than a microampere with 12 volts across it if you could measure it at all. You probably found the leakage in the silicon diode too low to be measured with your test set up.

This proves that the reverse resistance of the germanium diode is less than the silicon diode. It also accounts for the lower barrier voltage of the germanium diode (0.3 volts) compared to 0.7 volts for the silicon diode.

# DIODE CONSTRUCTION

A PN junction diode cannot be formed by simply pressing a section of P-type material against a section of N-type material. It simply is not possible to obtain the intimate contact that is necessary between the two materials by using a method of this type. In actual practice, PN junctions are formed by using procedures that are somewhat complex although the basic principles involved are really quite simple. We will now briefly examine several of these basic construction techniques so that you will have a better understanding of how these junctions are actually created. Then we will see how these junctions are placed in suitable packages or containers to obtain a finished diode.

## Grown Junctions

One of the earliest and most popular construction techniques is known as the *grown* method. With this method, PN junctions are actually grown by placing an intrinsic semiconductor and a P-type impurity into a quartz crucible (container) which is heated until the two materials melt. A small semiconductor crystal called a seed is then lowered into the molten mixture. The seed crystal is then slowly rotated and withdrawn from the molten mixture at a rate which will allow the molten material to cling to the seed. The molten material which clings to the seed crystal eventually cools and hardens and also assumes that same crystalline characteristics as the seed from which it appears to grow.

If the molten mixture contained only a P-type impurity as previously indicated, the grown crystal would be a P-type semiconductor. However, as the seed crystal is withdrawn, the molten mixture is alternately doped with N-type and P-type impurities. This means that N-type and P-type layers are created within the semiconductor crystal as it is grown. The resulting crystal is then sliced into many PN junctions.

## Alloyed Junctions

Junctions may also be constructed by a technique known as the *alloyed* method. This method is extremely simple as it is performed by simply placing a small pellet of indium on the N-type semiconductor crystal as shown in Figure 2-16. The pellet and semiconductor are then heated until the pellet melts into and partially fuses with the semiconductor crystal as shown. Since indium is a trivalent impurity, the region where the two materials combine becomes a P-type semiconductor material as shown in Figure 2-16. Once the heat is removed the two materials recrystallize and a solid PN junction is formed.

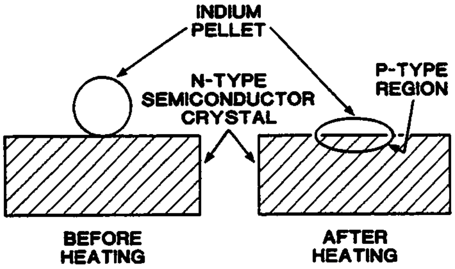


Figure 2-16
Alloyed PN Junction.

## Diffused Junctions

One of the most popular and preferred techniques used to construct PN junctions is referred to as the *diffusion* method. With this method, a thin section of N-type or P-type semiconductor material sometimes called a wafer or die is exposed to an impurity element which is in a gaseous or vaporized state. The operation occurs at a very high temperature and the impurity atoms penetrate or diffuse through the exposed surfaces of the wafer. The basic diffusion process is shown in Figure 2-17. First a mask is placed over a N-type semiconductor crystal. Then the mask and the crystal are exposed to a gaseous trivalent impurity as shown. The mask has an opening which allows the gaseous impurity to strike a limited portion of the crystal's surface. The impurity element therefore diffuses into the N-type crystal and forms a P-type semiconductor material. The result is a PN junction as shown. The depth of the diffused impurity is controlled by regulating the time of exposure and the temperature of the materials. This diffusion process can also be performed by eliminating the mask and simply letting the entire semiconductor crystal be exposed to the gaseous impurity. However, it then becomes necessary to trim away the undesired portions of the crystal so that only the desired PN junction remains.

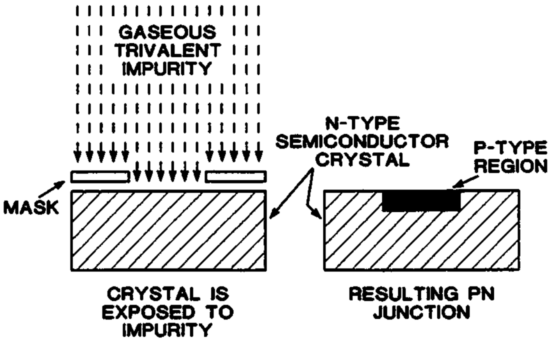


Figure 2-17
Diffused PN Junction.

# Diode Packaging

Once the PN junction is formed by using one of the three basic construction methods just described or by using similar or related techniques, the completed junction must be installed in a suitable container or package to produce the finished diode. This package protects the PN junction from environmental and mechanical stresses and at the same time provides an efficient means of connecting the PN junction to other components or parts. The type of container or package selected for each PN junction is determined by the particular purpose or application of each diode. For example, if a PN junction is designed to handle large currents, it must be mounted in a package that will aid in this purpose. In other words the package must be designed so that it will keep the junction from overheating due to the high current flow. This type of PN junction diode is often constructed as shown in Figure 2-18. Notice that the semiconductor wafer (PN junction) is sandwiched between two metal discs and then attached to a heavy metal base which has a threaded bolt or stud. This stud serves as an electrical connection to one side of the junction, usually the N-side or cathode. A heavy metal lug provides an electrical connection to the other side of the junction. This lug is attached to the disc on the P-side or anode through an S-shaped wire. This flexible wire is bonded to the lug and to the disc and it insures that the PN junction is not subjected to the same stress and strain that might be placed on the metal lug. The metal lug is held in place by the glass top and is therefore, insulated from the metal base and the metal case. The diode is used by simply bolting it to a metal surface which will help to dissipate the heat from the device. Since the threaded stud is connected to the N-side of the junction, the stud serves as the cathode connection and any metal chassis, frame, or plate that comes in contact with the stud is electrically connected to the cathode. The metal lug connects to the P-side of the junction and therefore serves as the anode connection. A wire or a component lead may be soldered to this lug to complete the installation of the diode.

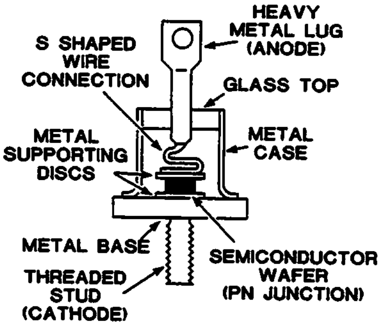


Figure 2-18
Cut-away view of a typical stud-mounted PN junction diode.

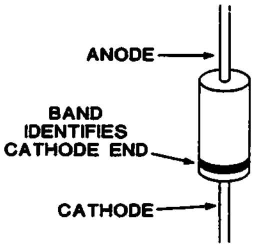


Figure 2-19
Typical semiconductor diode package.

The most common package for housing a PN junction diode is shown in Figure 2-19. The diodes contained in such packages are usually designed to handle small currents of 3 amperes or less. Most semiconductor diodes fall into this category. The semiconductor wafer (PN junction) is contained within a small cylindrical glass or epoxy case. The leads reach into the case to make contact with the wafer. The diode has axial leads and therefore can be installed in a variety of ways. In most cases the leads are simply soldered to a printed circuit board. The cathode end of the diode is usually identified by the band as shown. Sometimes a diode symbol is printed on the case to identify the leads.

The exact package and lead sizes depend on the type of diode, its application, and current handling capability. Most diodes are about 1/4" long and 1/8" in diameter.

## Self Test Review

35. A PN junction diode can be formed by tightly pressing N-type and P-type materials together.

    a. True
    b. False

36. The grown method of PN junction formation utilizes a ______________ crystal which is dipped into a molten mixture of semiconductor materials and then withdrawn.

37. The construction technique which allows a pellet of idium to be fused to a semiconductor crystal is referred to as the ______________ method.

38. The construction techniques which allows a gaseous impurity element to penetrate a semiconductor wafer is referred to as the ______________ method.

39. A ______________ is sometimes used to control the location of the N-type or P-type region that is formed when a semiconductor wafer is subjected to a gaseous impurity.

40. A completed PN junction diode consists of a PN junction which is mounted in a suitable ______________.

41. The cathode end of a diode may be identified by a ______________ at one end of the device.

42. Diodes which use a heavy mounting stud or plate are usually designed to conduct relatively large forward currents.

    a. True
    b. False

## Answers

35. (b) False
36. seed
37. alloyed
38. diffusion
39. mask
40. package
41. band
42. (a) True

# EXPERIMENT 2

# Testing Semiconductor Diodes

*OBJECTIVE:*

1. *To demonstrate a practical method of testing semiconductor diodes.*
2. *To determine if the diodes in your parts pack are shorted or open.*

## Introduction

Semiconductor diodes can be tested by a variety of methods. The oscilloscope, curve tracers, and specially constructed transistor testers can be used to determine the characteristics and conditions of a diode. However, in most cases diodes are tested primarily to see if they are simply good or bad, rather than to determine their detailed specifications. For that reason, a simple "go/no go" test method is preferred in most situations. A quick and easy diode test can be made with any standard ohmmeter. This experiment describes the procedure and the theory used when testing diodes with an ohmmeter. **Read the entire procedure before performing this experiment.**

## Material Required

Ohmmeter
1 — Germanium diode (56-602)
1 — Silicon diode (56-28)

NOTE: Many digital and electronic meters now provide a diode check mode, indicated by a diode symbol. In which, a constant current source instead of a constant voltage source is sometimes provided. A silicon diode will indicate between 0.5 and 0.7 $V_F$. A germanium diode should indicate between 0.2 and 0.4 $V_F$. In some cases, a diode is indicated, but the meter will still measure ohms. This meter position is designed to measure the forward and reverse resistances of a diode.

## Procedure

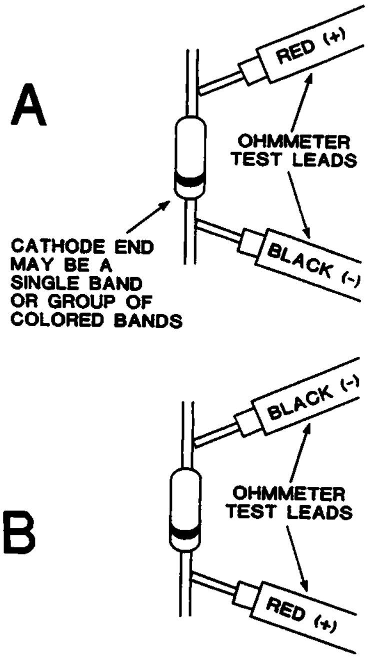


Figure 2-20
Typical test connections for measuring forward and reverse diode resistances.

1. In this first step, you will use your ohmmeter to test the germanium diode (56-602) supplied with this course. Set your ohmmeter to its lowest resistance range (R × 1) and connect the test leads across the diode, as shown in Figure 2-20A. As indicated in the figure, the red test lead is positive and the black test lead is negative, however, the test lead polarity may be different on analog meters. Make sure that the negative, or common, lead is connected to the cathode end of the diode. Your meter is now measuring the forward resistance ($R_F$) of the diode. If your meter is an analog type, set the range so that the meter indication is near the center or upper part of the scale. Record the forward resistance in the space provided in Figure 2-21.

2. Now set your ohmmeter to the R × 10k or R × 100k range and connect the meter's leads across the diode as shown in Figure 2-20B. Note that these connections are the reverse of those in Step 1. Your ohmmeter is now measuring the reverse resistance ($R_R$) of the diode. If necessary (analog meters) set your meter to another range to obtain a meter indication that is near the center or upper part of the scale. Record the indicated reverse resistance in the appropriate space provided in the table shown in Figure 2-21.

3. Repeat Steps 1 and 2 for your silicon diode (56-28) and record the diode's forward resistance ($R_F$) and reverse resistance ($R_R$) in the table provided in Figure 2-21.

| DIODE | $R_F$ | $R_R$ | $R_R/R_F$ |
|---|---|---|---|
| Ge | | | |
| Si | | | |

Figure 2-21
Table for recording the forward and reverse diode resistances.

4. Using the forward and reverse resistance values in Figure 2-21, calculate the ratio of the reverse to forward resistance $R_R/R_F$ and record it in the table.

## Discussion

Your measurements should show that the germanium diode has a very low forward resistance that is between a few ohms and several hundred ohms and a very high reverse resistance that is greater than 100 kilohms, and possibly as high as several megohms. The forward resistance of the silicon diode should be slightly higher than the forward resistance of the germanium diode, but again, should not exceed a value of several hundred ohms. The reverse resistance of the silicon diode should be much higher than the reverse resistance of the germanium diode. In fact, your meter probably indicated that the silicon diode has an infinite reverse resistance.

The reverse to forward resistance ratio RR/RF gives a relative indication of how well the diode passes current in one direction and blocks it in the other. The higher this ratio the better the diode. In Step 4 you should have found the ratio of the silicon diode to be much higher than that of the germanium. This significant difference is primarily the result of a high leakage current in the germanium device.

When using an ohmmeter to check the forward and reverse resistance of germanium and silicon diodes, it is possible to obtain only relative indications of resistance in each direction. Also, the resistance in either direction will change if you set your meter to a different resistance range. Figure 2-22A shows the equivalent circuit formed when an ohmmeter is used to check the forward resistance of a diode. Notice that the negative (–) lead of the ohmmeter, has an internal resistance as shown. The battery causes a small forward current to flow through the diode and also through its internal resistance. When different ranges are used, various internal resistances are switched into the circuit and the current through the diode changes. This causes the resistance of the diode to change and accounts for the various forward resistance readings that are obtained from one range the next and also from one ohmmeter to the next. A similar situation also occurs when the ohmmeter is used to measure the reverse resistance of the diode as shown in Figure 2-22B.

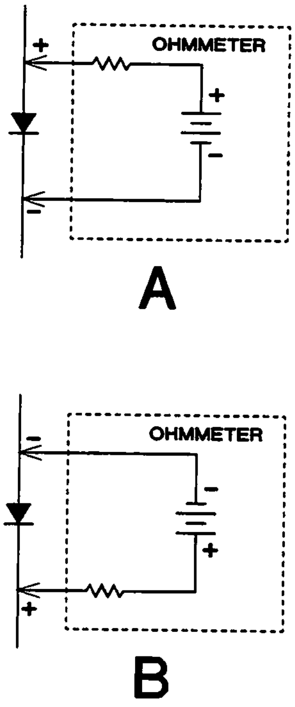


Figure 2-22
Equivalent ohmmeter circuit.

The ohmmeter circuit in Figure 2-22 shows the negative terminal of the internal battery connected to the negative, black or common lead. The positive battery terminal is connected to the positive lead. This is the arrangement in most ohmmeters since it permits you to quickly determine the polarity of the internal voltage at the ohmmeter leads. While this is a convenient and also obvious connection, some ohmmeters use just the opposite arrangement where the positive terminal of the internal battery connects to the negative or common lead. If the experiment worked as described above, then your ohmmeter is connected as shown in Figure 2-22. However, if your forward and reverse resistance measurements were obtained with connections opposite of those indicated in Figure 2-20, then you have the ohmmeter connection just described. Accurate diode tests can be made with any type of ohmmeter regardless of the internal circuitry. It is simply convenient to know what your meter circuitry is like.

Although ohmmeter readings are only relative, they do indicate when a diode meets its basic requirements by providing a high resistance in one direction and a low resistance in the opposite direction. If a diode has an extremely low resistance in both directions it usually has an internal short and is defective. Also, when a diode has an extremely high resistance in both directions it is considered to be open and therefore defective.

# DIODE APPLICATIONS

PN junction diodes are used extensively in the electronics industry. In some applications they are used simply because they are able to conduct current in the forward direction and block current in the reverse direction. However, in other applications they are used because of additional special characteristics that they possess. Many of the applications which utilize these special characteristics are discussed in the next two Units. At this time we are primarily concerned with the diode's forward conduction and reverse blocking capabilities. This fundamental ability of the diode to pass current in only one direction is utilized in a process known as rectification. We will now briefly examine this rectification process which is often considered to be the most important application of diodes in electronics.

## Rectification

Rectification is the process of converting alternating current into direct current. since diodes are used to accomplish this process they are often referred to as rectifiers. Rectification is a process that is required in most types of electronic equipment that must operate from alternating current. The primary power source for most electronic equipment is the ac power line. However, most electronic circuits usually require direct current for proper operation. The rectification process is therefore necessary to convert the available power (usually 110 to 220 volts, ac) into usable dc power which may be as low as several volts or as high as several thousand volts. The rectifier diode therefore does not operate alone in the rectification process. In fact a number of additional components are usually required. When assembled, these components form what is known as a power supply.

Figure 2-23 shows the basic rectification process. Notice that the input ac voltage is first applied to transformer which either steps up the input voltage or reduces it to a lower value as required. The ac voltage at the secondary of the transformer is then applied to a diode and resistor which are in series. Although

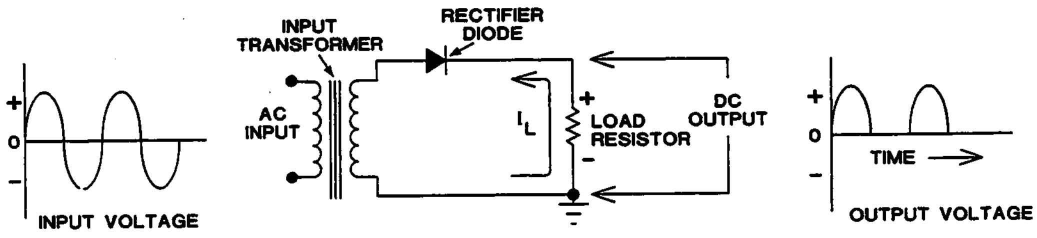


Figure 2-23
Typical rectifier circuit.

this resistor is shown as an individual component it actually represents the load which is the resistance of the electronic circuits which require dc power for operation. Since the diode is in series with the load resistance, it will allow current to flow through the diode and load resistance during each alternation of the input voltage that causes the diode to be forward biased. In other words, the diode will conduct on every other alternation of the input ac signal and therefore allow the current to flow through the load resistor in pulses. The pulsating load current ($I_L$) causes a pulsating output voltage to be developed across the load resistor as shown. This pulsating dc voltage is positive with respect to circuit ground because of the direction of the current through the resistor. The output voltage could be made negative with respect to ground by simply reversing the diode connections so that the direction of the current through the load resistor is reversed.

The pulsating dc output obtained from the rectifier circuit in Figure 2-23 does not usually provide an adequate operating voltage for most circuits. Most circuits require pure or continuous dc power for proper operation. This means that it is usually necessary to add extra components to the rectifier circuit to filter or smooth out the pulsations so that a constant output is obtained. The most commonly used filter component is a large capacitor connected across the load. The capacitor charges up to the peak value of the dc pulse during the time the diode conducts. During the next half cycle, the diode is reverse biased and does not conduct. During this time the capacitor discharges into the load to help maintain a continuous current flow. If the capacitor is large enough, the dc voltage across the load will be essentially constant.

The circuit shown in Figure 2-23 is known as a half wave rectifier since it conducts on only one half cycle of the ac input. Many other rectifier circuits are available to produce ac to dc conversion while providing other benefits as well.

## Other Applications for Diodes

While rectification is by far the most common use of diodes, there are many electronic circuits that take advantage of the unilateral characteristics of the diode. For example, the rectifier circuit in Figure 2-23 is also used to demodulate amplitude modulated radio signals. This is the circuit that removes the audio signal from the carrier. Used in this way, the circuit is called a detector. Other diode detector circuits are used to demodulate FM and SSB signals and a variety of other signals.

Another application for diodes is waveshaping. This is the process of converting one type of signal into another. For example, a sine wave may be readily converted into a square wave by using diodes.

Diodes are also widely used for protection purposes. In this application diodes are used to prevent damage to circuits or other components.

Diodes are often used as switches in automatic switching applications. Their polarity sensitive nature makes them easy to control in such situations.

Another application for diodes is in digital logic circuits. These are circuits that make decisions based on various input signals. Generally these circuits generate an output signal that is the result of a combination of input signals.

These are only a few examples of diode applications. In the next two units you will see further applications for special types of semiconductor diodes.

## Self Test Review

43. A diode used to convert ac into pulsating dc is called a _______________.

44. The component most often used to filter the pulsating dc into pure dc is a _______________.

45. A transformer, a rectifier diode and a filter are combined to form a circuit called a _______________ _______________.

46. Four other common applications for diodes are:

    a.
    b.
    c.
    d.

## Answers

43. rectifier
44. capacitor
45. power supply
46. a. detection or demodulation
    b. waveshaping
    c. protection
    d. logic circuits

# UNIT SUMMARY

Semiconductor diodes are produced by combining N-type and P-type semiconductor materials, forming a PN junction. The combination of these two materials allows free electrons and holes to combine until a depletion region, containing no majority carriers, is formed in the immediate vicinity of the PN junction.

The combination of electrons and holes causes positive charges to develop in the N-type material and negative charges to develop in the P-type material. These charges result from the ions that are left behind. Consequently, a barrier voltage develops across the junction. Eventually, the barrier voltage increases to a level that prevents further combining of electrons and holes, limiting the size of the depletion region. Germanium diodes have a typical barrier voltage of 0.3 volts. Similarly, a typical silicon diode barrier voltage is 0.7 volts.

When an external voltage that can overcome the barrier voltage is applied to the diode, the diode is forward biased, provided that the polarity of the voltage is correct. To forward bias a PN junction diode with a battery, the negative terminal must be connected to the N-type (cathode) end of the diode and the positive battery terminal must be connected to the P-type (anode) end of the diode. A forward-biased diode conducts by means of majority carriers. The voltage drop across the PN junction of a forward-biased diode is equal to the barrier voltage.

Reversing the polarity of the voltage applied to the diode, so that the anode is negative with respect to the cathode, reverse biases the diode. This is because the PN junction barrier voltage increases until it is equal and opposite to that of the applied voltage. This effectively cuts off almost all current through the diode. The small amount of reverse, or leakage, current through the diode consists of minority carriers and is usually insignificant. Conduction of a forward-biased diode is much heavier than that of a reverse-biased diode. Therefore, the forward resistance of a diode is much lower than its reverse resistance.

If the reverse bias voltage reaches a high enough value, the diode breaks down and conducts a relatively high reverse current. The voltage that causes this heavy reverse current is called the breakdown voltage of the diode. Heavy reverse current produces heat that can cause permanent damage to the diode unless a resistor is connected in series to limit reverse current to a level that is safe for the diode. The maximum reverse voltage that a diode can withstand without damage is called Peak Inverse Voltage (PIV). Diode manufacturers usually specify the PIV for each diode type.

Most diode characteristics are affected by temperature. Reverse leakage current and forward voltage drop are affected the most. In a typical diode, a ten-degree centigrade rise in temperature causes reverse current to nearly double. The forward voltage drop is inversely proportional to temperature; as temperature rises, the voltage drop decreases.

One method of forming a PN junction is by alternately doping intrinsic semiconductor material as a crystal is being grown from a seed. In another technique, an impurity is fused to material that is oppositely doped. However, the most common procedure used in the production of a PN junction is to inject an impurity into a previously doped semiconductor, forming a section of oppositely doped material.

After the PN junction is formed, electrical leads are added before the device is sealed in a protective package. The sealed package protects the PN junction from the environment, while the leads provide the means to connect the diode to other components in a circuit. Several package styles are used for housing diodes, depending on the intended application. All of these packages provide a means of cathode and anode identification.

The diode's ability to conduct well in one direction while effectively blocking current in the other direction has many applications in electronic circuits. One application is as a rectifier for converting ac to dc. Other applications include detection, wave shaping, logic, and circuit protection.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. When a PN junction is formed, the area around the junction loses its majority carriers. Therefore, the area is referred to as the:

   A. vacuum region
   B. barrier zone
   C. depletion voltage
   D. depletion region

2. The positive and negative ions on each side of the PN junction create an internal voltage which is called a:

   A. barrier voltage
   B. breakdown voltage
   C. depletion region
   D. depletion voltage

3. When a PN junction diode is reverse-biased, the device will:

   A. conduct a relatively high forward current.
   B. conduct a small leakage current.
   C. conduct a high reverse current.
   D. not conduct until the external bias voltage is greater than 0.7 volts.

4. The majority carriers in a forward-biased PN junction diode are forced to:

   A. combine at the outer edges of the diode.
   B. move away from the junction and widen the depletion area.
   C. leave the diode and combine within the battery.
   D. combine at the junction.

5. A diode's V-I characteristic curve shows how the diodes:

   A. forward current varies with temperature.
   B. barrier voltage varies with temperature.
   C. forward and reverse currents vary when the forward and reverse voltages are changed.
   D. depletion area varies with forward voltage.

6. The reverse voltage that causes the reverse current in a diode to suddenly increase is referred to as the diode's:

A. reverse bias voltage.
B. breakdown voltage.
C. peak inverse voltage.
D. forward saturation voltage.

7. The forward voltage across a conducting silicon diode will:

A. always be less than 0.7 volts.
B. be equal to the external bias voltage.
C. have a typical value at 0.3 volts.
D. be higher than the forward voltage across a germanium diode that is operating under similar conditions.

8. When a PN junction is formed by allowing a pellet of indium to fuse or melt into an N-type semiconductor crystal, the construction is referred to as the:

A. alloyed method.
B. grown method.
C. diffusion method.
D. depletion method.

9. The bar or rectangle portion of the diode symbol represents the:

A. P section of the diode.
B. anode end of the diode.
C. cathode end of the diode.
D. type of material used to construct the diode.

10. When a diode is used to convert alternating current into direct current it is commonly referred to as a:

A. alternator
B. rectifier
C. commutator
D. regulator

11. The current flowing in the circuit of Figure 2-24 is ____________, and the voltage drop across the diode is ____________.

A. practically zero; 6 Volts
B. 0.7 mA; 5.3 volts
C. 5.3 mA; 0.7 volts
D. 6 mA; 0.3 volts

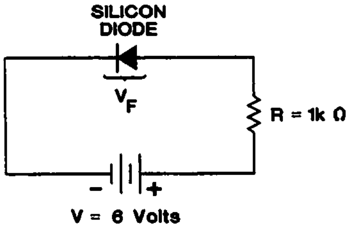


Figure 2-24
Circuit for exam question 11.

12. The current flowing in the circuit of Figure 2-25 is ____________, and the voltage across the diode is ____________.

    A. practically zero; 12 volts.
    B. 3.9 mA; 11.7 volts.
    C. 0.1 mA; 0.7 volts.
    D. 4 mA; 0.3 volts.

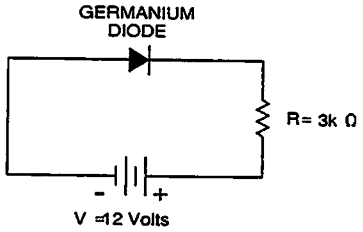


Figure 2-25
Circuit for exam question 12

13. The most widely used method of constructing a semiconductor diode is by the:

    A. alloyed method.
    B. grown method.
    C. diffused method.
    D. Ohmmeter

14. The instrument most widely used to test semiconductor diodes is a:

    A. voltmeter.
    B. ammeter.
    C. Oscilloscope.
    D. ohmmeter.

15. Temperature variations cause which of the following diode characteristics to vary the most.

    A. Forward voltage
    B. Forward Current
    C. Leakage Current
    D. PIV

# EXAMINATION ANSWERS

1. D — The area near the junction is called the depletion region because it is effectively depleted of majority carriers.
2. A — The internal difference potential is called a barrier voltage because it forms a potential barrier which limits the size of the depletion region.
3. B — A reverse-biased PN junction diode allows only a small reverse or leakage current to flow as long as its breakdown voltage is not exceeded.
4. D — The majority carriers (holes and electrons) combine at the junction and neutralize each other. This action supports forward current flow through the diode.
5. C — The V-I (voltage-current) characteristic curve is extremely helpful when analyzing the overall operation of a diode.
6. B — When the breakdown voltage is reached, the diode effectively breaks down and allows a relatively high reverse current to flow.
7. D — Silicon diodes have a higher internal barrier voltage than germanium diodes. Therefore a higher forward voltage is required to cause the silicon diode to conduct forward current. The forward drop is typically .7 volts for a silicon diode.
8. A — Since the indium pellet actually fuses or alloys to the semiconductor crystal, the process is called the alloyed method.
9. C — The bar in the diode symbol represents the N section of the diode which is commonly referred to as the cathode.
10. B — Since the process of converting alternating current to direct current is known as rectification, diodes used for this purpose are referred to simply as rectifiers.

11. C — The external bias battery is connected to properly forward bias the diode. The diode conducts and current flows in the circuit. The current is:

$$I = \frac{V - V_F}{R} = \frac{6 - .7}{1000} = \frac{5.3}{1000} = .0053 \text{ amps} = 5.3 \text{ MA}$$

The forward voltage drop of a silicon diode is 0.7 volts.

12. A — The external bias battery is connected to reverse bias the diode. Little or no current flows except for leakage. The reverse voltage across the diode is the supply voltage 12 volts.

13. C — The diffusion method is the most widely used process of making PN junction diodes.

14. D — An ohmmeter is commonly used to test diodes although special diode and transistor testers are available.

15. C — Temperature variations affect all diode characteristics, but the leakage current, particularly in germanium diodes, is affected most.

## *Unit 3*

# THE ZENER DIODE

# CONTENTS

Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-3

Unit Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-3

Unit Activity Guide . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-4

Zener Diode Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-5

    Voltage-Current Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-6

    Zener Voltage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-7

Power Dissipation in Zener Diodes . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-8

    Power — Temperature Curves . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-8

    Derating Factor . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-9

Current Limitations in Zener Diodes . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-11

Effects of Temperature on Zener Voltage . . . . . . . . . . . . . . . . . . . . . . . Page 3-11

Temperature Compensated Zener Diodes . . . . . . . . . . . . . . . . . . . . . . . Page 3-12

Zener Diode Impedance . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-14

Zener Diode Packages . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-15

Voltage Regulation with Zener Diodes . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-15

    The Basic Zener Diode Regulator . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-16

    Designing a Zener Regulator Circuit . . . . . . . . . . . . . . . . . . . . . . . . Page 3-18

Experiment 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-23

Unit Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-30

Unit Examination . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-32

Examination Answers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Page 3-35

# INTRODUCTION

You are now ready to examine an electronic component that is closely related to the PN junction diode that you examined in the previous unit. This device is commonly referred to as a zener diode and it is widely used throughout the electronics industry. Because the zener diode is an extremely important electronic component, you must study this unit of instruction very carefully. By doing so, you will become familiar with another important solid-state component and at the same time expand your knowledge of semiconductor diodes in general.

Be sure to examine your unit objectives closely and then follow the instructions provided in your unit activity guide to be sure that you perform all of the necessary steps as you proceed through this unit.

# UNIT OBJECTIVES

When you have completed this unit, you should be able to:

1. Describe the forward and reverse current-voltage characteristics of a typical zener diode.
2. Understand the relationship between temperature and zener diode power dissipation.
3. Know how to determine a zener diode's maximum safe operating current.
4. Know how to use a diode's zener impedance to determine the amount of change that can occur in the diode's zener voltage.
5. Describe how the zener diode is used to provide voltage regulation.
6. Design a zener regulator circuit.

# UNIT ACTIVITY GUIDE

| | Completion Time |
|---|---|
| ☐ Read "Zener Diode Characteristics" and "Power Dissipation in Zener Diodes." | |
| ☐ Answer Self Test Review Questions 1-9 | |
| ☐ Read "Current Limitations in Zener Diodes," "Effects Of Temperature On Zener Voltage," and "Temperature Compensated Zener Diodes." | |
| ☐ Answer Self Test Review Questions 10-18 | |
| ☐ Read "Zener Diode Impedance," "Zener Diode Packages," and "Voltage Regulation With Zener Diodes." | |
| ☐ Answer Self Test Review Questions 19-31 | |
| ☐ Perform Experiment 3 on "Zener Diodes." | |
| ☐ Read the Unit Summary | |
| ☐ Complete the Unit Examination | |
| ☐ Check the Examination Answers | |

# ZENER DIODE CHARACTERISTICS

In the previous unit you learned that an ordinary PN junction diode breaks down and conducts a relatively high reverse current when it is subjected to a sufficiently high reverse bias voltage. This high reverse current occurs because the high reverse voltage is capable of tearing valence electrons away form their parent atoms and increasing the number of minority carriers in the N and P sections of the diode. The reverse voltage which causes an ordinary PN junction diode to break down is the diode's breakdown voltage.

Ordinary PN junction diodes can be damaged if they are subjected to their respective breakdown voltages. This is because the high reverse currents can produce more heat than the diodes can safely dissipate. However, special diodes are constructed which can operate at voltages that equal or exceed their breakdown voltage ratings. These special diodes are called *zener diodes.*

We will now examine the exact relationship that exists between the current flowing through a zener diode and the voltage across the device. We will consider the action that takes place when the zener diode is forward-biased and reverse-biased, but we will be primarily concerned with the action that takes place at the point where breakdown occurs. Then we will see how a zener diode is rated according to its breakdown voltage.

Although very brief, this discussion on zener diode characteristics is very important. Pay particular attention to the new terms and symbols that are introduced. You will use these terms and symbols throughout this unit.

# Voltage-Current Characteristics

A typical zener diode V-I (voltage-current) characteristic curve is shown in Figure 3-1. Notice that the overall forward and reverse characteristics of the zener diode are similar to those of an ordinary junction diode. The primary difference is simply that the zener diode is specifically designed to operate with a reverse bias voltage that is high enough to cause the device to breakdown and conduct a high reverse current. As shown in Figure 3-1, the zener diode's reverse current remains at a very low value until the reverse voltage is increased to a value that is sufficient to cause the diode to breakdown. Then the reverse current through the diode increases at an extremely rapid rate as the reverse voltage increases beyond the breakdown point. The V-I curve therefore shows that beyond the breakdown point, a very large change in reverse current is accompanied by only a very small change in reverse voltage. This action occurs because the resistance of the diode drops considerably as its reverse voltage is increased beyond the breakdown point. Once the breakdown point is exceeded the diode is said to be operating in its *zener breakdown region* or simply its *zener region*. The current that flows through the diode while it is operating in its zener region is often referred to as the *zener current* and may be represented by the symbol $I_Z$.

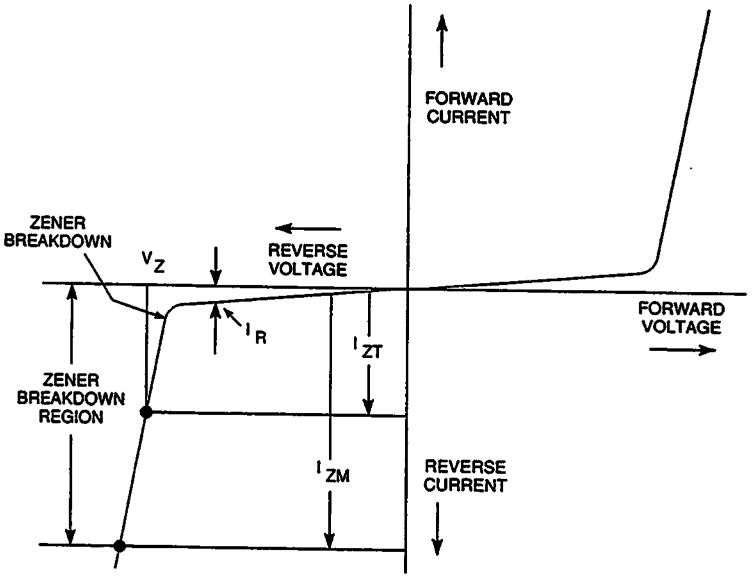


Figure 3-1
Typical VI Characteristic
Curve for a Zener Diode.

If you examine Figure 3-1 closely you will see that breakdown (also called zener breakdown) does not occur instantaneously. The curve is rounded near the breakdown point. This curved or rounded portion is often called the *knee* of the curve. When a zener diode has a knee with a very sharp edge, the diode enters the breakdown region very quickly. However, when the knee is more rounded, the breakdown region is entered at a slower rate. The importance of this consideration will be explained later in this discussion.

## Zener Voltage

The breakdown voltage of a zener diode is determined by the resistivity of the diode which in turn can be controlled by the various doping techniques that are used to form the device. A zener diode is manufactured to have a specific breakdown voltage rating which is often referred to as the diode's *zener voltage* and is designated as $V_Z$. Typical $V_Z$ values may vary from several volts to several hundred volts. For example, some of the popular low voltage units have ratings of 3.3, 4.7, 5.1, 5.6, 6.2, and 9.1 volts; however, additional voltage ratings are also available. It is important to realize that when a zener diode is rated as having a specific zener voltage ($V_Z$), the rated voltage does not represent the reverse voltage that is required to initially cause the diode to breakdown. The rated zener voltage is a nominal value that represents the reverse voltage across the diode when the zener current is at some specified value called the *zener test current* ($I_{ZT}$). The V-I curve in Figure 3-1 shows the relative zener voltage ($V_Z$) and zener test current ($I_{ZT}$) values for a typical zener diode. Notice that these values are located within the zener breakdown region of the curve. The zener test current ($I_{ZT}$) simply represents a typical value of reverse current that is always less than the maximum reverse current that the diode can safely handle.

Like resistors and capacitors, zener diodes cannot be produced with zener breakdown voltages that are always exactly equal to a specified value. Therefore, it is necessary to specify minimum and maximum breakdown voltage limits for each device. This is done by specifying a breakdown voltage tolerance for each type of diode that is manufactured. The standard zener breakdown voltage (zener voltage) tolerances are ±20 percent, ±10 percent, and ±5 percent, however specially manufactured zener diodes are also available in ±1 percent tolerance. For example, a 6.8 volt ±10 percent zener diode will have a zener voltage somewhere in the 6.12 to 7.48 volt range.

# POWER DISSIPATION IN ZENER DIODES

Manufacturers of zener diodes also specify the maximum power dissipation of each device. Some devices are rated at only several hundred milliwatts while others are rated as high as 50 watts. However, some of the most popular and widely used devices have relatively low ratings of 400 milliwatts, 500 milliwatts, and 1 watt. A zener diode's power dissipation rating is given for a specific operating temperature. Often the power rating is given for a temperature of 25° centigrade, 50° centigrade, or 75° centigrade. However, the actual power that a zener diode can safely dissipate will decrease if the temperature increases above this specified level or increase if the temperature decreases below the specified level. Also, if the diode has axial leads, its power rating is often specified for a specific level. Also, if the diode has axial leads, its power rating is often specified for a specific lead length or various ratings are given for various lead lengths. This is because a diode's ability to dissipate power increases as its leads are shortened. The shorter leads (when appropriately soldered in an electronic circuit) are more effective in conducting heat away from the diode's PN junction.

## Power — Temperature Curves

To simplify the relationship between a zener diode's maximum power rating, its temperature, and its lead length, a power-temperature derating curve is often supplied with each type of diode manufactured. A typical curve for a diode that has a power dissipation rating of 500 milliwatts at a temperature of 70° with lead lengths of 3/8" is shown in Figure 3-2. Notice that three curves are shown for three different lead lengths of 1/8", 3/8", and 1". The specified power rating of 500 milliwatts occurs only when the lead length is equal to 3/8" and the temperature is equal to 75° centigrade as shown. If the temperature increases above or decreases below 75° centigrade, the power rating decreases below or increases above 500 milliwatts respectively. Also notice that the shorter lead length (1/8") allows the diode to dissipate more power over the same temperature spread while the longer lead length (1") reduces the overall power rating of the device. The curve also shows that the power rating of the device is effectively reduced to zero at 200° centigrade. As you examine Figure 3-2 keep in mind that the actual temperature of the diodes leads are indicated and not just the ambient or surrounding air temperature which is sometimes shown in less specific power temperature curves. The diode's leads are also assumed to be soldered to a suitable circuit board or component which can serve as a heat sink to drain away the heat produced by the device.

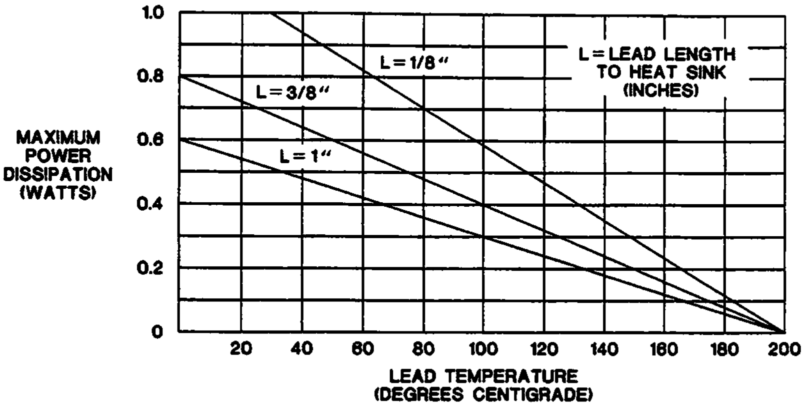


Figure 3-2

Power-Temperature Derating Curve.

# Derating Factor

If the zener diode is a high power type that is contained in a metal case or package that must be stud-mounted, a power-temperature derating curve may be provided which shows the diode's power rating for various case temperatures. However, in some cases the manufacturer will simply give a specific power rating for a zener diode (regardless of type) at a specific lead, case, or ambient temperature and then specify what is known as a *derating factor*. The derating factor is usually given in milliwatts per degree centigrade, and can be used to determine the power rating of the diode at temperatures that are different (usually higher) than the one specified. For example, a typical zener diode may have a derating factor of 6 milliwatts per degree centigrade. This simply means that the diode's power rating decreases 6 milliwatts for every degree centigrade increase in temperature.

## Self Test Review

1. Zener diodes are designed to safely operate within their zener breakdown regions.

   a. True
   b. False

2. The curved portion of the V-I curve near the point where zener breakdown occurs, is referred to as the __________ of the curve.

3. The current which flows through a zener diode that is operating in its zener breakdown region is referred to as the diode's ___________ current.

4. Since it is impossible to manufacture zener diodes with breakdown voltages that are exactly equal to a specified value, it is necessary to indicate the minimum and maximum breakdown voltage limits by specifying a zener voltage ___________ for each component.

5. A zener diode's power dissipation rating is usually given for a specific operating _____________.

6. In general, the actual power that a zener diode can safely dissipate will decrease if temperature _____________.

7. A diode's ability to dissipate power is increased when its leads are ____________.

8. The relationship between a zener diode's maximum power rating, its temperature, and its lead length is often expressed graphically in the form of a ________ ________ ________ curve.

9. A zener diode's power rating for various temperatures can also be determined by using a ________ ________ which is usually given in milliwatts per degree centigrade.

## Answers

1. (a) True
2. knee
3. zener
4. tolerance
5. temperature
6. increases
7. shortened
8. power-temperature derating
9. derating factor

# CURRENT LIMITATIONS IN ZENER DIODES

The maximum reverse current that can flow through a zener diode without exceeding the diode's power dissipation rating is commonly referred to as the *maximum zener current* and is represented in the V-I curve in Figure 3-1 by the symbol $I_{ZM}$. The $I_{ZM}$ value of a zener diode is often specified by the manufacturer of the device. However, if $I_{ZM}$ is not specified, it can be determined by simply dividing the power dissipation rating of the diode by its breakdown voltage (zener voltage) rating or stated mathematically:

$$I_{ZM} = \frac{\text{power rating}}{\text{zener voltage}}$$

However, it is best to play it safe and use the maximum limit of zener voltage in your calculations. For example, suppose you have a zener diode that is rated at 10 watts, and the diode has a zener voltage of 5.1 volts at ±10 percent tolerance. The maximum voltage limit would be equal to 5.1 volts plus 10 percent of 5.1 volts or 5.1 + 0.51 which is equal to 5.61 volts. The maximum zener current would therefore be equal to

$$I_{ZM} = \frac{10}{5.61} = 1.78 \text{ amperes}$$

The V-I curve in Figure 3-1 also shows that a small reverse or leakage current ($I_R$) flows through the zener diode before the breakdown point is reached. Since the zener diode is normally used in its zener breakdown region, this current is usually not important. However, there are certain applications of zener diodes which require an absolute minimum leakage current before the breakdown point is reached. Therefore manufacturers often specify the $I_R$ value of a zener diode at a certain reverse voltage that is less than the zener voltage ($V_Z$) (often 80 percent of $V_Z$).

# EFFECTS OF TEMPERATURE ON ZENER VOLTAGE

Zener diodes also have other characteristics that must be considered in certain applications. For example, a diode's zener voltage will vary slightly as temperature changes. The amount of voltage change that takes place is usually expressed as a percentage of zener voltage ($V_Z$) change for each degree centigrade rise in temperature and is referred to as the *zener voltage temperature coefficient.* Zener diodes that have a zener breakdown voltage of 5 volts or more, usually have positive zener voltage temperature coefficients. This simply means that their breakdown voltages increase as temperature increases. However, most diodes that have breakdown voltages that are below approximately 4 volts usually have a negative zener voltage temperature coefficient. This means that the breakdown voltage will decrease with an increase in temperature. When the breakdown voltages are between approximately 4 and 5 volts, the zener voltage temperature coefficient may be either positive or negative. For example, a zener diode with a zener breakdown voltage of 3.9 volts might have a zener voltage temperature coefficient of −.025 percent per degree centigrade. This means that the diode's zener voltage will decrease .025 percent (or approximately .001 volt) for each degree centigrade rise in temperature.

## TEMPERATURE COMPENSATED ZENER DIODES

Special zener diodes are constructed which are temperature compensated so that their zener voltage ratings remain almost constant with changes in temperature. These special diodes are commonly referred to as temperature compensated zener diodes or voltage reference diodes. A temperature compensated diode is formed by connecting a zener in series with an ordinary PN junction diode. However, the two are connected back-to-back so that the junction diode will be forward-biased while the zener diode is reverse-biased. The zener diode usually has a zener voltage rating that is greater than 5 volts and therefore has a positive temperature coefficient. However, the forward-biased diode will have a forward voltage drop of approximately 0.6 or 0.7 volts and a negative temperature coefficient. By carefully selecting the two devices so that their temperature coefficients are equal and opposite, the voltage changes effectively cancel out. Furthermore, the voltage drops across the two devices must be summed to obtain the overall voltage rating of the temperature compensated device. For example, when a 5.6 volt zener diode is connected in series with a junction diode that has a forward voltage drop of 0.6 volts, a 6.2 volt temperature compensated zener diode is produced. In some cases more than one junction diode may be used to obtain the necessary compensation. Typical temperature compensated zener diodes will have temperature coefficients that range from .01 percent per degree centigrade to .0005 percent per degree centigrade. However, the optimum temperature stability usually occurs at or near a specific operating current which is normally specified by the manufacturer.

# Self Test Review

10. The maximum reverse current that can flow through a zener diode without exceeding the diode's power dissipation rating is referred to as the diode's __________ __________ __________.

11. The maximum safe value of reverse current that can flow through a zener diode can be calculated by dividing the diodes power rating by its __________ __________.

12. Manufacturers often specify the __________ current that flows through a zener diode before its breakdown point is reached.

13. A diode's zener voltage temperature coefficient is usually expressed as a percentage of zener voltage change for each degree centigrade rise in __________.

14. A diode that has a zener breakdown voltage of 9.1 volts will have a __________ zener voltage temperature coefficient.

15. A diode that has a zener breakdown voltage of 3.3 volts will usually have a __________ zener voltage temperature coefficient.

16. When a diode's breakdown voltage decreases as temperature increases, the device has a __________ zener voltage temperature coefficient.

17. Temperature compensated zener diodes are formed by connecting a zener diode in series with one or more __________ __________ __________.

18. In a temperature compensated zener diode the zener diode portion of a device usually has a __________ temperature coefficient.

## Answers

10. maximum zener current
11. zener voltage
12. reverse or leakage
13. temperature
14. positive
15. negative
16. negative
17. PN junction diodes
18. positive

# ZENER DIODE IMPEDANCE

Another important characteristic that should be considered when examining any type of zener diode is the diode's *zener impedance* ($Z_{ZT}$). This is determined by varying the zener current above and below the specified zener test current ($I_{ZT}$) value and then observing the corresponding change in zener voltage ($V_Z$) as shown in Figure 3-3. The zener impedance is equal to the change in zener voltage ($\Delta V_Z$) divided by the change in zener current ($\Delta I_Z$) and will vary considerably from one diode to the next. Some zener diodes with low zener voltage ratings will have a $Z_{ZT}$ of only a few ohms. In general, the lower the zener impedance the greater the slope of the curve in the zener breakdown region. A low $Z_{ZT}$ therefore indicates that the zener voltage changes only slightly with changes in zener current. An ideal zener diode would not change its breakdown voltage as zener current varies and would therefore have a zener impedance of zero ohms. The zener impedance of a diode is also useful in determining the changes in zener voltage which can occur when the diode is used at zener currents which are higher or lower than $I_{ZT}$. It is a simple matter to calculate this change in zener voltage by using a known value of zener impedance and the change or deviation in zener current. When expressed mathematically, the change in zener voltage is equal to

$$\Delta V_Z = \Delta I_Z \times Z_{ZT}$$

The same technique used to determine $Z_{ZT}$ may also be used to determine the impedance at the knee of the curve near the point where breakdown occurs. The impedance at the knee of the curve is commonly referred to as the *zener knee impedance* ($Z_{ZK}$). The zener knee impedance provides an indication of the slope or sharpness of the knee of the curve. Manufacturers of zener diodes will usually specify both the zener impedance ($Z_{ZT}$) and the zener knee impedance ($Z_{ZK}$) for each device.

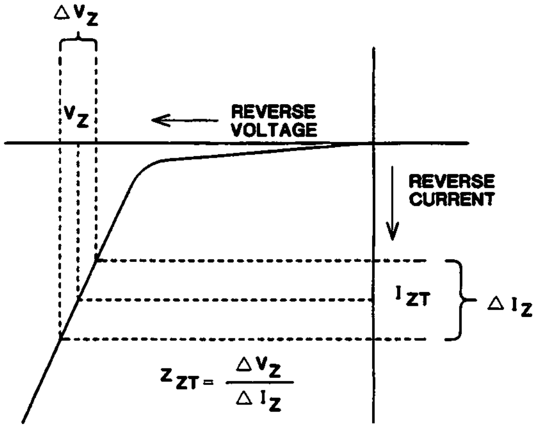


Figure 3-3
The Basic Method Used to Determine Zener Impedance ($Z_{ZT}$).

# ZENER DIODE PACKAGES

Basically, zener diodes are packaged in the same manner as the ordinary PN junction diodes described in the previous unit. The low power devices usually have axial leads and are mounted in either glass or epoxy cases while the high power units are usually stud mounted and are contained in metal cases. A typical low power zener diode is shown in Figure 3-4A and the commonly used zener diode symbol is shown in Figure 3-4B. A band is used to identify the cathode end of the zener diode as shown, therefore, the zener diode resembles an ordinary PN junction diode in appearance. The zener diode symbol is also similar to the ordinary diode symbol. The only difference is that the cathode end of the diode is represented by a zig-zag or Z shaped bar instead of a straight bar.

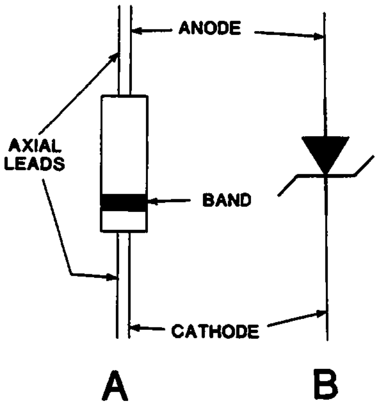


Figure 3-4
A Typical Zener Diode and its Schematic Symbol.

# VOLTAGE REGULATION WITH ZENER DIODES

Although the zener diode may be used to perform number of important function it is perhaps most widely used in applications where it is continually reverse-biased so that it operates constantly within its zener breakdown region. Under these conditions, the zener diode is effectively used to provide voltage stabilization or regulation.

Voltage regulation is often required because most solid-state circuits require a fixed or constant dc power supply voltage for proper operation. If this dc voltage changes significantly from the required value, improper operation will usually result. If an ac line operated power supply is used (one that is not regulated), the dc output voltage will vary if the power line voltage changes or if the load resistance connected to the power supply changes. However, by using a zener diode regulator circuit., it is possible to compensate for these changes and maintain a constant dc output voltage.

# The Basic Zener Diode Regulator

A typical zener diode regulator circuit is shown in Figure 3-5. Notice that the diode is connected in series with a resistor, and an unregulated (fluctuating or charging) dc input voltage is applied to these two components. The input voltage is connected so that the zener diode is reversed-biased as shown in Figure 3-5. The series resistor allows enough current to flow through the diode so that the device operates within its zener breakdown region. In order for this circuit to function properly, the input dc voltage must be higher than the zener breakdown voltage rating of the diode.

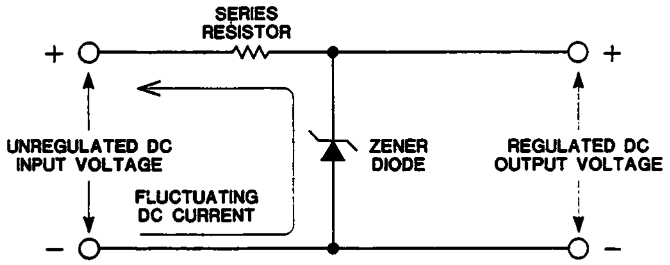


**Figure 3-5**
A Typical Zener Diode Voltage Regulator Circuit.

The voltage across the zener diode will then be equal to the diodes zener voltage rating and the voltage across the resistor will be equal to the difference between the diode's zener voltage ($V_Z$) and the input dc voltage.

The input dc voltage shown in Figure 3-5 is unregulated or in other words it is not held to a constant dc value. This voltage may periodically increase above or decrease below its specified value. Therefore, the dc current flowing through the zener diode and the series resistor will fluctuate accordingly. However, the zener diode is operating within its zener voltage region and a wide range of zener currents can flow through the diode, but the zener voltage changes only slightly. This tendency of the diode to oppose any change in voltage results because the resistance of the diode drops when zener current increases as described earlier. Since the diode's voltage remains almost constant as the input voltage varies, the change in input voltage almost completely appears across the series resistor which has a fixed or constant resistance. You must remember that these two components are in series and the sum of their voltage drops must always be equal to the input voltage.

The voltage across the zener diode is used as the output voltage for the regulator circuit. The output voltage is therefore equal to the diodes zener voltage and since this voltage is held to a somewhat constant value, it is referred to as a regulated voltage. The output voltage of the regulator circuit can be changed by simply using a zener diode with a different zener voltage rating and selecting a series resistor that will allow the diode to operate within its zener breakdown region.

Voltage regulator circuits like the one shown in Figure 3-5 are used in many types of electronic equipment to provide constant operating voltages for various circuits. However, these circuits require not only operating voltage but various operating currents as well. Furthermore, there are many cases where operating currents continually vary because circuit impedances are not constant. When designing a regulator circuit it is therefore necessary to consider the specific current or the range of currents that the regulator must supply as well as the output voltage that the regulator must provide.

When a circuit, a lamp, or some other device is connected to the output of the regulator circuit in Figure 3-5, the regulator must supply current as well as voltage to the external device (commonly called a load). This situation is shown in Figure 3-6. Notice that the load resistor ($R_L$) requires a specific load current ($I_L$) which is determined by its resistance and the output voltage. The current through the zener diode ($I_Z$) combines with $I_L$ and flows through the series resistor ($R_S$). The value of $R_S$ must therefore be chosen so the $I_Z$ remains at a sufficient level to keep the diode within its breakdown region and at the same time allow the required value of $I_L$ to flow through the load.

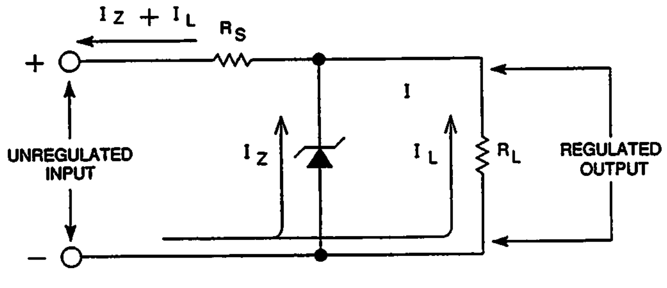


Figure 3-6

Loading Affects in a
Zener Diode Voltage Regulator.

If the load current increases or decreases because $R_L$ effectively decreases or increases in value, it might appear that the output voltage would change accordingly. In other words even if the input voltage remains constant there is a tendency for the output voltage to increase or decrease if $R_L$ increases or decreases in value respectively. If $R_L$ and $R_S$ were in series and the zener diode was not installed in the circuit, this action would definitely result. However, the zener diode prevents this from happening. When $R_L$ increases in value, $I_L$ decreases and the voltage across $R_L$ tries to increase. However, the zener diode opposes this change by conducting more zener current ($I_Z$) so that the total current ($I_Z + I_L$) flowing through $R_S$ remains essentially constant (assuming that the input voltage remains constant). The same voltage is therefore maintained across $R_S$ and also across the parallel branch (the zener diode and $R_L$). Also, whenever $I_L$ increases, $I_Z$ decreases by approximately the same amount to again hold $I_Z + I_L$ almost constant and therefore maintain an essentially constant output voltage. In this way the zener diode regulator is able to maintain a relatively constant output voltage even though changes in output current occur. The circuit therefore regulates for changes in output current as well as for changes in input voltage.

## Designing a Zener Regulator Circuit

To further illustrate the use of the zener diode in a typical voltage regulator circuit we will now design a typical regulator using some basic design rules. We will begin by assuming that certain input and output conditions must exist and we will calculate the necessary component values. First we will assume that the unregulated input voltage varies from a minimum value of 9 volts to a maximum value of 12 volts. The regulated output voltage will be equal to 5.1 volts which is a standard zener voltage rating. The output load current will vary from 0 to 30 milliamperes. To satisfy these conditions we will of course need a zener diode with a zener voltage rating of 5.1 volts; however, we must also use a series resistor ($R_S$) that will allow the diode to operate properly with the above mentioned changes in input voltage and output current. The required value of $R_S$ can be quickly determined by using the following equation.

$$R_S = \frac{V_{in\,(min)} - V_Z}{1.1\, I_{L\,(max)}}$$

This equation states that $R_S$ is equal to $V_{in\ (min)}$ (the minimum value of input voltage) minus $V_Z$ (the zener voltage) divided by 1.1 $I_{L\ (max)}$ (1.1 times the maximum value of load current). By subtracting $V_Z$ from $V_{in\ (min)}$ we obtain the minimum voltage that will be dropped across $R_S$. This voltage must then be divided by 1.1 $I_{L\ (max)}$ to determine the required value of $R_S$. Multiplying 1.1 times $I_{L\ (max)}$ is the same as increasing the value of $I_{L\ (max)}$ by 10 percent. This additional current is used as a safety factor to insure that the zener diode's current does not drop below the level needed to keep the diode operating within its zener breakdown region when the unregulated input voltage is at its minimum value. You must remember that the total current is equal to $I_Z + I_L$ and when $I_L$ is at a maximum value, $I_Z$ must decrease to a minimum value. By specifying a maximum load current that is 10 percent higher than required, we are insuring that this additional current (which is not used by the load) will always flow through the diode. In most cases 10 percent is considered to be sufficient.

When we substitute actual values in the equation shown above we find that the series resistance must be equal to

$$R_S = \frac{9 - 5.1}{1.1\,(.03)} = 118 \text{ ohms}$$

Since this is not a readily available standard resistance value, we must select the next lower value of resistance available which is 100 ohms. A lower value is selected to insure that $I_Z$ will not drop below the level necessary to keep the diode operating within its zener breakdown region.

Now we must determine the maximum power that the zener diode will be required to dissipate. The maximum power dissipated by the zener diode can be easily calculated by using the following equation.

$$P_{Z\ (max)} = V_Z \left[ \frac{V_{in\ (max)} - V_Z}{R_S} - I_{L\ (min)} \right]$$

This equation simply states that $P_{Z\ (max)}$ (the maximum power dissipated by the zener diode) is equal to $V_Z$ (the zener voltage) times the maximum value of zener current. However, the maximum zener current is determined by finding the maximum current through $R_S$ and then subtracting $I_{L\ (min)}$ (the minimum load current) from this maximum current. The maximum current through $R_S$ is found by subtracting $V_Z$ (the zener voltage) from $V_{in\ (max)}$ (the maximum input voltage) and dividing the difference by $R_S$.

When we substitute actual values into this equation we find that the zener diode must dissipate a maximum power of

$$P_{Z\ (max)} = 5.1\left(\frac{12 - 5.1}{100} - 0\right) = 0.352 \text{ watts}$$

However, in our circuit we would use a zener diode with a power dissipation rating that is slightly higher than 0.352 watts (352 milliwatts) to provide an additional safety margin. If the diode is to be used where the ambient temperature is near 25° centigrade (normal room temperature) we could select a diode that is rated at either 400 milliwatts or 500 milliwatts at this temperature.

Throughout all of our calculations we have assumed that the zener voltage $V_Z$ remains essentially constant. This simplifies our calculations and in most cases it is a practical assumption. In cases where the actual changes in $V_Z$ (usually only several tenths of a volt) must be considered, the diode's zener impedance (specified by the manufacturer) is helpful in determining these changes as zener current varies. It is also important to select a diode with a zener voltage tolerance that is suitable for the particular application. A ±20 percent tolerance might be acceptable in a non critical application, but a lower tolerance of ±5 percent or ±1 percent may be needed where the output voltage must be very close to the specified value.

# Self Test Review

19. A diode's zener impedance is determined by dividing a change in __________ __________ by a corresponding change in zener current.

20. An ideal zener diode would have a zener impedance of ___________ ohms.

21. The impedance at the knee of a zener diode's V-I curve is referred to as the diode's __________ __________ _________.

22. Low power zener diodes usually have ___________ leads.

23. High power zener diodes are usually __________ mounted.

24. The zener diode symbol is similar to the PN junction diode symbol but has a __________ shaped bar instead of a straight bar.

25. A zener diode is often used in applications where it is continually __________-biased.

26. When used in a voltage regulator circuit the zener diode must be subjected to an input dc voltage that is higher than the zener breakdown voltage rating of the diode.

    a. True
    b. False

27. The output voltage produced by the voltage regulator circuit in Figure 3-5 is equal to the zener diode's __________ __________ rating.

28. The voltage across the series resistor in Figure Figure 3-5 remains constant as the unregulated input voltage varies.

    a. True
    b. False

29. The circuit shown in Figure 3-6 provides a relatively constant output voltage even though changes in input voltage and changes in __________ __________ occur.

30. The zener current ($I_Z$) in Figure 3-6 should __________ when the load current ($I_L$) increases so that the output voltage will remain constant.

31. Design a zener regulator whose output voltage is 12 volts. The load current is fixed at 40 mA but the input voltage varies over a 15 to 20 volt range. Calculate $R_S$ and the power dissipation of the zener.

## Answers

19. zener voltage
20 zero
21. zener knee impedance
22. axial
23. stud
24. zig-zag or Z
25. reverse
26. (a) True
27. Zener voltage
28. (b) False
29. output current or load resistance
30. decrease
31.

$$R_S = \frac{V_{in\,(min)} - V_Z}{1.1\, I_{L\,(max)}} = \frac{15-12}{1.1\,(.04)} = \frac{3}{.044} = 68.18 \text{ ohms}$$

A standard 68 ohm resistor could be used.

$$P_Z = V_Z \left( \frac{V_{in\,(max)} - V_Z}{R_S} - I_{L\,(min)} \right)$$

$$P_Z = 12 \left( \frac{20-12}{68} - .04 \right)$$

$$P_Z = 12 \left( \frac{8}{68} - .04 \right) = 12\,(.1176 - .04)$$

$$P_Z = 12\,(.0776) = .9312 \text{ watt}$$

Use a 1 watt zener

# EXPERIMENT 3

# Zener Diodes

*OBJECTIVE:*

1. *Demonstrate the characteristics of a zener diode.*
2. *Investigate a zener diode's application as a voltage regulator.*

## Introduction

In this experiment, you will demonstrate the operational characteristics of a typical low power zener diode. You will measure its performance under forward and reverse bias conditions. Then, you will use this diode in a voltage regulator circuit. You will demonstrate the ability of the circuit to compensate for input voltage and load variations. Read the entire procedure before performing this experiment.

## Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — 47 ohm, resistor (yellow, violet, black, gold)
1 — 100 ohm, 2-watt resistor (brown, black, brown, gold)
1 — 220 ohm, resistor (red, red, brown, gold)
1 — 470 ohm, resistor (yellow, violet, brown, gold)
1 — 1 kilohm, resistor (brown, black, red, gold)
1 — 1N5231B or 1N751 Zener diode, 5.1 volts ±10 percent, 500 milliwatts (Part No. 56-16)

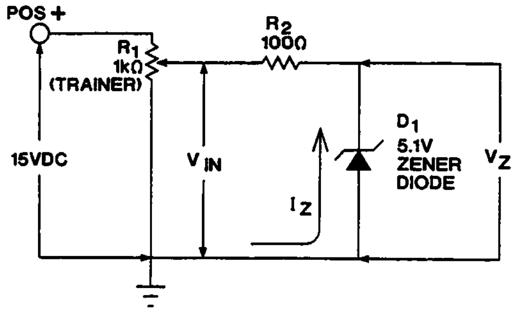


**Figure 3-7**
Experimental Circuit.

## Procedure

1. Construct the circuit shown in Figure 3-7, using the 1 kilohm potentiometer on your Trainer, a 100 ohm resistor, and the zener diode, as shown in Figure 3-8. Do not turn on Trainer power at this time.

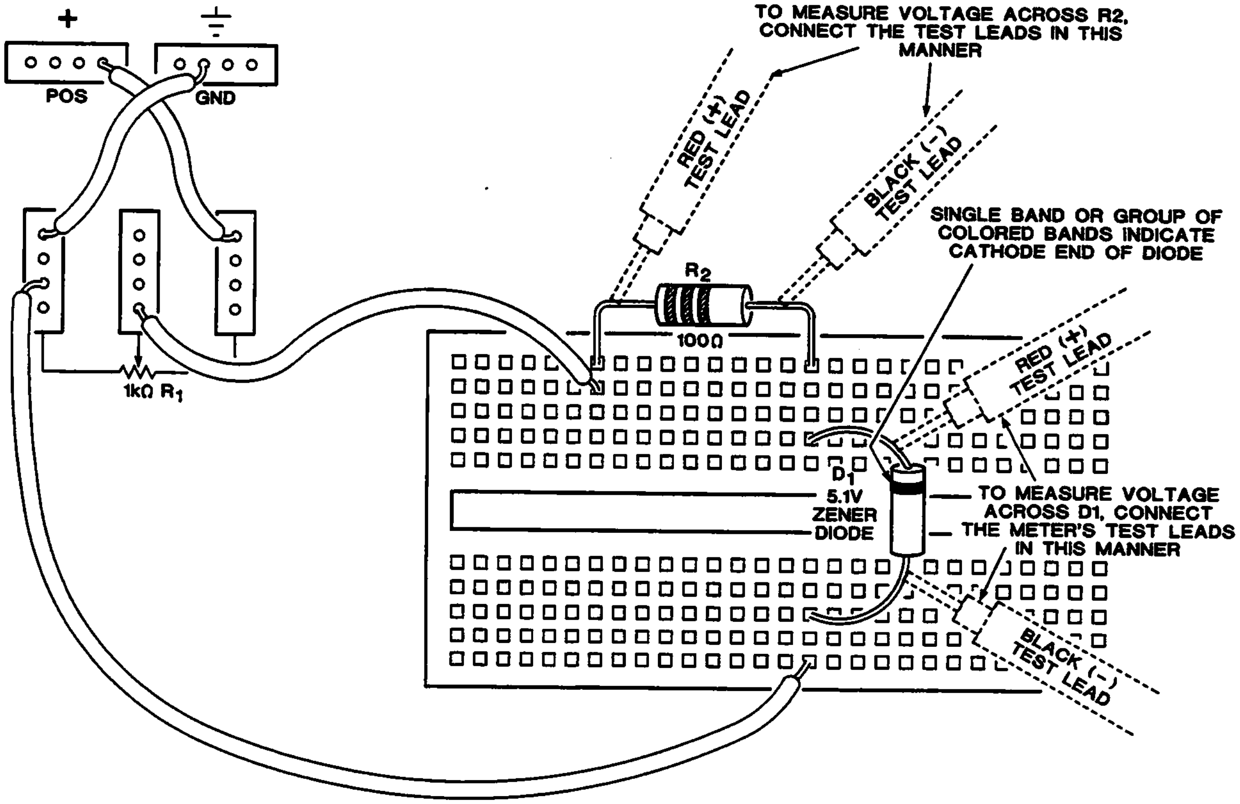

+
POS
GND
TO MEASURE VOLTAGE ACROSS R2, CONNECT THE TEST LEADS IN THIS MANNER
RED (+) TEST LEAD
BLACK (-) TEST LEAD
SINGLE BAND OR GROUP OF COLORED BANDS INDICATE CATHODE END OF DIODE
R2
100Ω
1kΩ R1
RED (+) TEST LEAD
D1
5.1V
ZENER
DIODE
TO MEASURE VOLTAGE ACROSS D1, CONNECT THE METER'S TEST LEADS IN THIS MANNER
BLACK (-) TEST LEAD

**Figure 3-8**
Wiring diagram
for Experimental Circuit.

2. Turn the positive (+) voltage control fully clockwise so that the full power supply voltage will be applied to the 1 kilohm potentiometer. You will be using the 1 kilohm potentiometer (designated as $R_1$) to control the voltage applied to the zener diode ($D_1$) and its 100 ohm series resistor ($R_2$).

   Observe the polarity of the voltage applied to $D_1$ in Figure 3-7. Is $D_1$ forward or reverse-biased?______________

3. Adjust the 1 kilohm potentiometer $R_1$ for maximum resistance, then turn on your Trainer.

4. Connect your voltmeter across $D_1$ as indicated in Figure 3-8. Then, adjust $R_1$ slowly and observe the increase in voltage across $D_1$ as indicated on your meter. Continue turning $R_1$ until the voltage across $D_1$ stops increasing at a rapid rate and effectively remains constant. At this point stop turning $R_1$ and, as accurately as you can, measure the voltage across $D_1$. Record this voltage in the DIODE VOLTAGE column (upper space) of the table shown in Figure 3-9.

| POSITION OF $R_1$ | DIODE VOLTAGE | DIODE CURRENT | INPUT VOLTAGE |
|---|---|---|---|
| (REFER TO TEXT) | | | |
| MINIMUM RESISTANCE | | | |

**Figure 3-9**
Chart for Recording Zener Diode Regulator Current and Voltages.

5. Next, use your voltmeter to measure the voltage across $R_2$. Then use the voltage across $R_2$ and the resistance of $R_2$ (100 ohms) to calculate the current flowing in the circuit. Your calculated value of current represents the amount of current that is now flowing through resistor $R_2$ and diode $D_1$. Record this current in the DIODE CURRENT column (upper space) in Figure 3-9.

6. Measure the input voltage ($V_{in}$) between the potentiometer wiper terminal and ground as indicated in Figure 3-7. Record your value in the INPUT VOLTAGE column (upper space) of the table shown in Figure 3-9.

7. Now you will make several voltage measurements which must be performed quickly to avoid overheating the diode. First, measure the voltage across $D_1$ as you turn potentiometer $R_1$ fully clockwise. Note the voltage across $D_1$ with $R_1$ fully clockwise and record this voltage in the DIODE VOLTAGE column (lower space) of the table shown in Figure 3-9. Next, measure the input voltage and record this value in Figure 3-9. Then measure the voltage across $R_2$. The voltage across $R_2$ is now equal to ______________ volts. Now turn off your Trainer.

8. Use the voltage across $R_2$ (recorded above) and the resistance of $R_2$ to calculate the current flowing through $R_2$. This calculated value of current represents the current that flows through $R_2$ and $D_1$. Record this current in the DIODE CURRENT column (lower space) in Figure 3-9.

9. Reverse the diode leads and turn $R_1$ fully counterclockwise. Then turn on the Trainer. Connect your voltmeter between the wiper terminal of potentiometer $R_1$ and ground, adjust the input voltage ($V_{in}$) to 5 volts.

10. Measure the voltage across the diode with your voltmeter. The diode voltage is now equal to ________ volts. Is the diode forward or reverse-biased? __________. Turn off your Trainer and read the following discussion.

## Discussion

In this part of the experiment, you subjected the zener diode to a variable reverse voltage, and recorded the voltage across the device and the current through the device at two important points. The diode was **reverse biased** since its cathode was positive with respect to its anode.

The diode voltage and current values that you recorded in the upper row of Figure 3-9 represent the point where the diode first enters the zener breakdown region. At this time, the zener voltage levels off at an almost constant value and the current just begins to increase at a rapid rate. You should have found that the diode voltage leveled off at a value of approximately 5.1 volts, since it has a zener voltage rating that is equal to this value. However, you may have measured a voltage that was several tenths of a volt higher or lower than this value because this diode has a zener voltage tolerance of ±10 percent. The current that you recorded should be quite low, since it represents the current through the diode just after the breakdown point is reached. You probably recorded a current that was between 2 and 5 milliamperes. The input voltage ($V_{in}$) at this time was probably about 5.5 to 6 volts.

Next, you set the variable input voltage to its maximum value by adjusting the 1 kilohm potentiometer for minimum resistance. At this time, the current flowing through $R_2$ and $D_1$ was much higher. As you increased the input voltage you should have found that the voltage across $D_1$ increased only slightly (perhaps only one or two tenths of a volt). the diode voltages that you recorded in Figure 3-9 should, therefore, be very close to one another. However, the diode current that you recorded for maximum input voltage should be much higher than the current that you measured initially. You should have obtained a diode current that was between 80 and 100 milliamperes. At this point, the zener is operating at the limit of its power dissipation rating. The input voltage at this time should have been above 15 volts.

A direct comparison of the diode currents and voltages in Figure 3-9 should show that the zener voltage changes only slightly while the current varies over a wide range while the diode is operating within its zener region. Comparing the input and zener voltages, you can see that as the input voltage varies, the zener voltage varies only a few tenths of a volt. This clearly demonstrates the value of the zener in compensating for input voltage variations.

In Steps 9 and 10 you reversed the diode leads. This causes the zener to be forward biased., When forward biased, the zener acts as any ordinary silicon PN junction diode. You should have measured a forward voltage drop in the 0.6 to 0.7 volt range.

## Procedure (continued)

11. You will now use your experimental circuit to supply a regulated output voltage of 5.1 volts to various resistive loads. When a load resistor is connected across the zener diode, the circuit effectively becomes a simple voltage regulator circuit. Your experimental circuit has the same component values as the regulator circuit described in the typical design procedure in the previous section. The circuit should, therefore, be capable of operating with input voltages between 9 and 12 volts and output load currents between 0 and 30 milliamperes. You will now test this circuit by setting the input voltage to its lower limit of 9 volts and observe the regulated output voltage for various load currents.

12. Reverse the diode leads to return the circuit to its original condition. Insure that the positive (+) voltage control is turned fully clockwise. Then, turn on your Trainer. Next, connect your voltmeter between the wiper of the potentiometer and ground (across $R_2$ and $D_1$) and adjust the 1 kilohm potentiometer ($R_1$) until the meter indicates that the voltage applied to $R_2$ and $D_1$ is exactly equal to 9 volts.

13. With no load resistor connected across diode $D_1$, (no load current) use your voltmeter to measure the voltage across $D_1$. Record the indicated voltage in the OUTPUT VOLTAGE column of the chart shown in Figure 3-10.

| LOAD RESISTANCE (OHMS) | APPROXIMATE LOAD CURRENT (MILLIAMPERES) | OUTPUT VOLTAGE (VOLTS) |
|---|---|---|
| OPEN (NO LOAD) | 0 | |
| 1000 | 5 | |
| 470 | 11 | |
| 220 | 23 | |
| 47 | 108 | |

**Figure 3-10**
Chart for Recording Output Voltage
of Regulator over a Range of Output Loads.

14. Now connect a 1000 ohm (1 kilohm) resistor across diode $D_1$. This resistor will serve as a load and will draw a load current of approximately 5 milliamperes at the rated output voltage of 5.1 volts. Measure the output voltage across $D_1$ at this time, and record the indicated voltage in the appropriate space in the OUTPUT VOLTAGE column of the chart shown in Figure 3-10.

15. Now, repeat Step 14 using load resistors with values of 470 ohms, 220 ohms, and 47 ohms. These resistors will draw load currents of approximately 11, 23, and 10-8 milliamperes, respectively, assuming the output voltage remains at 5.1 volts. After you connect each load resistor across $D_1$ and measure the voltage across $D_1$, record the indicated voltage in the space provided in Figure 3-11. Then, turn off your Trainer.

## Discussion

In this part of the experiment, you demonstrated the zener diode's ability to provide a regulated output voltage under various load conditions. You connected a series of resistors to the output of the regulator circuit to draw specific amounts of load current from the circuit and checked the output voltage in each case. You should have measured an output voltage of approximately 5.1 volts (±10 percent) across $D_1$, and this voltage should have remained essentially constant when the output was open (no load) or when the 1000 ohm, 470 ohm, or 220 ohm resistors were connected to the circuit. However, the output voltage should have dropped well below 5.1 volts when the 47 ohm load resistor was connected to the circuit. You probably measured a voltage of about 3 volts, or less. This decrease in output voltage should result because the 47 ohm resistor draws a load current of 108 milliamperes, at an output voltage of 5.1 volts. This situation cannot occur because the voltage regulator circuit is designed to provide a maximum output load current of 30 milliamperes when the input voltage is at a minimum value of 9 volts. Since all of your measurements were taken with the input voltage set to 9 volts, the 47 ohm resistor effectively overloaded the regulator circuit and caused the output voltage to drop.

Actually, when the 47 ohm load is connected, the load resistance on a zener regulator decreases and the load current increases proportionally. The zener current, therefore, decreases by a proportional amount. As long as the zener current remains high enough, the zener will continue to operate in its breakdown region and provide a relatively constant load voltage. If the load resistance becomes too low, and attempts to draw more current from the circuit than the series resistor will permit, the zener current drops to zero. In effect, the low load resistance takes all of the current. The effect is the same if the zener were removed entirely. In this case, the equivalent circuit simply becomes a voltage divider made up of the 100 ohm series resistor and the 47 ohm load. The resulting output voltage is determined by the voltage divider effect of these two resistors and input voltage. **This illustrates that the zener regulator is designed for a specific set of load conditions and if these conditions are exceeded, improper operation will result.**

# UNIT SUMMARY

PN junction diodes that are intended for use as rectifiers can be damaged if they are subjected to reverse voltages that equal or exceed their breakdown voltage ratings. This damage results from heat produced by the heavy reverse current through the device.

Zener diodes are designed to withstand the high reverse current that results when the diode's breakdown voltage is applied. After the breakdown voltage of the zener diode is exceeded, reverse current through the diode increases from an insignificant leakage value to a relatively high value. Once a zener diode is operating in its breakdown, or zener, region, an increase in current causes the internal resistance of the diode to decrease. Conversely, if current decreases, the diode internal resistance increases. As a result, the voltage drop across the zener diode remains essentially constant. Current through a zener diode operating in the zener region is called zener current. If zener current exceeds the specified maximum for a given diode, the diode can be permanently damaged.

You can determine the zener voltage rating ($V_z$) of a given zener diode by measuring the voltage drop across the diode with a specified zener test current ($I_{zt}$) through the diode. Zener test current is always greater than the reverse current that initiates zener breakdown, but less than the maximum current that the device can withstand.

Zener diodes are available for almost any zener voltage rating. However, it is impractical to manufacture a zener diode capable of dropping an exact specified voltage. Therefore, a certain amount of variation is expected. Consequently, manufacturers usually specify the amount of variation expected as a percentage of tolerance. For example, if a given zener diode is rated at 9.1 volts with a tolerance of ±20%, the actual voltage dropped by that diode may be any voltage in the range of approximately 7.3 volts to 10.9 volts.

Zener diode manufacturers also specify the maximum power that their products can safely dissipate at a given temperature. Diode power ratings increase or decrease with decreases or increases in temperature, respectively. In addition, you can increase the power rating of a zener diode with axial leads by shortening the length of the leads.

The most common application of zener diodes is as regulators in regulated dc electronic power supplies. When a zener diode is used as part of a properly designed regulator circuit, the diode maintains a constant (regulated) dc output voltage. In a typical regulator circuit, the diode is connected in series with a resistor to the unregulated output of the power supply. The load is then connected in parallel with the diode. Consequently, any variations in the unregulated output voltage caused by load current variations or line voltage changes are dropped across the resistor, while the voltage across the zener diode and the load remains constant. In addition, the sum of the zener and load currents remains fairly constant.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C or D) that you feel is most correct. When you have completed the entire examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. When a zener diode is operating in its zener breakdown region, a large change in zener current will produce

   A. a large change in zener voltage.
   B. a proportional change in zener voltage.
   C. a relatively small change in diode's reverse voltage.
   D. absolutely no change in the diode's reverse voltage.

2. A zener diode's power dissipation rating is usually given for specific

   A. operating voltage.
   B. zener test current.
   C. zener impedance.
   D. operating temperature.

3. In general, a properly installed zener diode can dissipate more power if its axial leads are

   A. shortened.
   B. curved.
   C. straight.
   D. lengthened.

4. The maximum reverse current that can blow through a zener diode without exceeding the diode's power dissipation rating is commonly designated as

   A. $I_{ZT}$
   B. $I_{ZM}$
   C. $I_{ZK}$
   D. $I_Z$

5. The exact amount of zener voltage change that occurs over a specific temperature range can be determined by using the diodes

   A. temperature compensation factor.
   B. zener impedance
   C. derating factor.
   D. zener voltage temperature coefficient.

6. When a zener diode has an extremely low zener impedance, its

   A. zener voltage varies only slightly with changes in zener current.
   B. zener voltage remains constant with changes in zener current.
   C. zener current varies only slightly while its zener voltage varies over a wide range.
   D. zener current remains constant while its zener voltage varies only a small amount.

7. A diode that has a zener voltage rating of 6.2 volts will usually have a

   A. high zener knee impedance.
   B. high power dissipation rating.
   C. positive zener voltage temperature coefficient.
   D. negative zener voltage temperature coefficient.

8. A diode's specified zener voltage ($V_Z$) is that voltage which is dropped across the diode when the current through the device is at a specific level called the

   A. maximum zener current.
   B. zener test current.
   C. minimum zener current.
   D. forward test current.

9. When used in a properly designed voltage regulator circuit, a zener diode is capable of maintaining a relatively constant output voltage

   A. as long as the output load current remains constant.
   B. as long as the input voltage remains constant.
   C. when the input voltage or the output load current changes.
   D. when only the output load current changes.

10. If the series resistance required by a voltage regulator is calculated to be 215 ohms, it would be desirable to use a standard resistance value of:

   A. 150 ohms.
   B. 180 ohms.
   C. 220 ohms.
   D. 270 ohms.

11. The maximum current that a 15 volt, 1 watt zener diode can handle without damage is

    A. 15 ma
    B. 66 ma
    C. 132 ma
    D. 1 amp.

12. To bias a zener for proper operation as a regulator

    A. the input voltage should be equal to the zener voltage.
    B. the load current should be less than the zener current.
    C. the anode should be positive with respect to the cathode.
    D. the cathode should be positive with respect to the anode.

13. A forward biased zener diode has a voltage drop equal to

    A. 0.3 volts
    B. 0.7 volts
    C. $V_Z$
    D. $V_Z$ + 0.7 volts

14. In a zener regulator circuit, an increase in input voltage will cause

    A. an increase in zener current.
    B. an increase in load current.
    C. an increase in load voltage.
    D. a decrease in zener current.

15. An increase in load resistance on a zener regulator will cause the zener current to

    A. decrease.
    B. increase.
    C. remain almost constant.

# EXAMINATION ANSWERS

1. C— The reverse voltage across the zener diode varies only slightly with changes in zener current as indicated by the diode's zener impedance ($Z_{ZT}$).

2. D— The power dissipation rating of a zener diode is greatly affected by temperature and is therefore specified at a particular ambient or lead temperature.

3. A— The shorter leads are more effective in conducting heat away from the diode if the device is properly soldered to a circuit board or some other component.

4. B— A diode's maximum reverse current is referred to as its maximum zener current which is designated as $I_{ZM}$.

5. D— The diode's zener voltage temperature coefficient shows the percentage of zener voltage change for each degree centigrade rise in temperature.

6. A— The lower the zener impedance, the smaller the change in zener voltage for a given change in zener current. An ideal diode would have a zener impedance of zero ohms and its zener voltage would remain constant.

7. C— Any zener diode that has a zener breakdown voltage of 5 volts or more will usually have a positive zener voltage temperature coefficient.

8. B— The zener test current ($I_{ZT}$) is a typical value of reverse current which is always less than the maximum zener current ($I_{ZM}$).

9. C— The voltage regulator is usually designed to compensate for specified ranges or input voltage and output load current.

10. B— You would use the next lower value of resistance available or 180 ohms.

11. B— $I_{ZM} = \frac{P_Z}{V_Z} = \frac{1}{15}$ = 0.0666 amperes or about 66 ma

12. D — To use the zener as a voltage regulator it must be reverse biased by making its cathode sufficiently positive with respect to its anode.

13. B — A forward-biased zener acts like any other silicon diode in that its forward drop will be approximately 0.7 volts.

14. A — Increasing the input voltage to a zener regulator will cause the zener current to increase. The load voltage and current will remain essentially constant.

15. B — An increase in load resistance will cause a decrease in load current. As a result, the zener current will increase.

## *Unit 4*

# SEMICONDUCTOR DIODES FOR SPECIAL APPLICATIONS

# CONTENTS

Introduction .......... Page 4-3

Unit Objectives .......... Page 4-3

Unit Activity Guide .......... Page 4-4

The Tunnel Diode .......... Page 4-5

  Voltage — Current Characteristics .......... Page 4-6

  What is Negative Resistance? .......... Page 4-7

  Construction .......... Page 4-8

  Applications .......... Page 4-8

The Varactor Diode .......... Page 4-12

  Electrical Characteristics .......... Page 4-12

  Operating Efficiency .......... Page 4-13

  Construction .......... Page 4-15

  Applications .......... Page 4-16

Experiment 4 .......... Page 4-17

High Frequency Diodes .......... Page 4-21

  PIN Diodes .......... Page 4-21

  Schottky Diodes .......... Page 4-22

  Gunn Effect Devices .......... Page 4-23

  IMPATT Diodes .......... Page 4-25

Unit Summary .......... Page 4-28

Unit Examination .......... Page 4-30

Examination Answers .......... Page 4-33

# INTRODUCTION

Now that you have examined the basic PN junction diode and the zener diode, it is time to examine some other types of diodes that have unique characteristics that make them suitable for special applications. Some of these special diodes are constructed in the same manner as the PN junction diodes previously described while others are formed by using entirely different construction techniques. Study this unit carefully as it will greatly expand your knowledge of semiconductor diodes and emphasize the important role that these devices play in electronics. Examine your unit objectives closely and carefully follow the instructions in your unit activity guide to be sure that you perform all of the necessary steps to complete this unit successfully.

# UNIT OBJECTIVES

When you have completed this unit on special semiconductor diodes you should be able to:

1. Locate the peak current, valley current, peak voltage, and valley voltage points on a tunnel diode's V-I curve.
2. Describe the negative resistance characteristic of a tunnel diode.
3. Recognize three commonly used tunnel diode symbols.
4. Describe how capacitance is produced within a varactor diode.
5. Describe how a varactor's internal capacitance and Q are affected by a change in operating voltage.
6. Recognize three commonly used varactor diode symbols.
7. Name four special diode types other than tunnel and varactor diodes, and state an application for each.
8. Demonstrate the capacitive characteristics of a varactor diode.

# UNIT ACTIVITY GUIDE

| | Completion Time |
|---|---|
| ☐ Read "The Tunnel Diode." | ____________ |
| ☐ Answer Self Test Review Questions: 1-10 | ____________ |
| ☐ Read "The Varactor Diode." | ____________ |
| ☐ Answer Self Test Review Questions: 11-22 | ____________ |
| ☐ Perform Experiment 4 on "Varactor Diodes." | ____________ |
| ☐ Read "High Frequency Diodes." | ____________ |
| ☐ Answer Self Test Review Questions: 23-31 | ____________ |
| ☐ Read the Unit Summary. | ____________ |
| ☐ Complete the Unit Examination. | ____________ |
| ☐ Check the Examination Answers | ____________ |

# THE TUNNEL DIODE

The ordinary PN junction diodes and zener diodes that you have already examined have lightly doped PN junctions and they have voltage-current characteristics that are quite similar. However, there is a special type of PN junction diode that is produced by using special construction techniques and heavy doping (a high concentration of impurities) to obtain characteristics that are radically different from those of ordinary PN junction or zener diodes. This device is commonly referred to as a *tunnel diode*.

Due to its heavily doped PN junction, the tunnel diode has a high internal barrier voltage and an extremely narrow depletion region. The device also has an extremely low reverse breakdown voltage (almost zero) and therefore conducts large currents when it is reversed-biased. The forward characteristics of the device are also unique. It would appear that the high barrier voltage would prevent forward current from flowing through the diode when it is subjected to low forward bias voltages; however, this is not the case. At low forward bias voltages, electrons are forced through the narrow depletion region at an extremely high velocity because of the high concentration of charges on each side of the junction. The electrons effectively appear to tunnel through the potential barrier (barrier voltage) as they move across the junction. Also, during this period of time when the tunneling action is occurring, a point is reached where an increase in forward voltage can actually cause a decrease in forward current through the diode.

## Voltage-Current Characteristics

A typical tunnel diode V-I characteristic curve is shown in Figure 4-1. As shown in this figure, the diode will conduct high reverse currents when it is subjected to reverse voltages. However, the diode's most important electrical characteristics occur when it is forward-biased. Notice that the forward current through the diode initially increases as the forward bias voltage is increased, but a point is soon reached where forward current stops increasing with an increase in forward voltage. The current that flows through the tunnel diode at this time is referred to as the *peak current* $I_p$ and the voltage across the diode is referred to as the *peak voltage* $V_p$. A further increase in forward voltage causes the forward current through the diode to decrease as shown. The current continues to decrease as forward voltage is increased until it reaches a minimum value which is referred to as the *valley current* $I_V$. At this time the voltage across the diode is referred to as the *valley voltage* $V_V$. If the forward voltage across the diode is increased still further, the diode's forward current will again increase. However, this time the current increases in the same way as it would in an ordinary PN junction diode that is subjected to an increasing forward bias voltage.

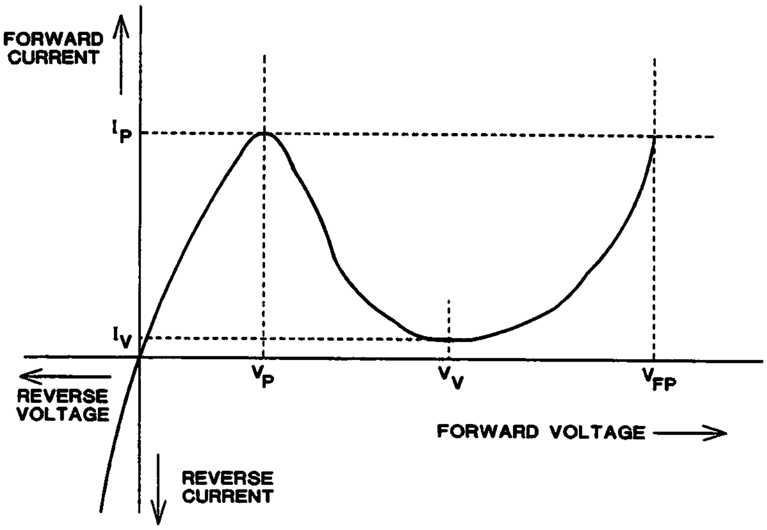


Figure 4-1 Typical tunnel diode V-I Characteristics

Between the peak and valley points on the V-I curve, the tunnel diode's forward current decreases as the forward voltage across the device is increased. This portion of the diode's V-I curve is therefore referred to as the *negative resistance* region. The tunneling action previously mentioned is reduced throughout this portion of the curve and it ceases when the $I_V$ value is reached. It is this negative resistance region that makes the tunnel diode an extremely useful electronic component.

# What is Negative Resistance?

Negative Resistance is that characteristic of an electronic component or circuit where the current through it decreases as the voltage across it increases and vice versa. The resistance of the component or circuit is not really negative in the true sense of the word. Its resistance is a positive value, but its effect defies Ohm's law as we know it. Ohm's law says that increasing the voltage across a resistance will result in an increase in the current through that resistance.

$$E = IR$$

Manufacturers of tunnel diodes usually specify the peak current $I_P$, peak voltage $V_P$, valley current $I_V$, and valley voltage $V_V$ values for each device. These values must be known in order to properly bias the tunnel diode within its negative resistance region. The peak current can be easily regulated regardless of the type of semiconductor material used. Most tunnel diodes designed to have low peak currents (often as low as 100 microamperes) but devices have been constructed that have peak currents as high as 10 amperes. The valley current is usually held to a low value with respect to the peak current so that a high peak-to-valley current ratio $I_P/I_V$ is maintained. The greater this ratio, the greater the operating current range within the negative resistance region. The peak voltage and valley voltage are determined by the type of semiconductor material used to construct the diode and for all practical purposes have fixed values. For example, a germanium tunnel diode will have typical $V_P$ and $V_V$ values of 55 and 350 millivolts respectively at 25° centigrade.

In addition to the $V_P$ and $V_V$ values, the manufacturer may also specify the *projected peak voltage* $V_{FP}$. This is the forward voltage at which the forward current rises to a point where it is again equal to $I_P$ as indicated in Figure 4-1. The $V_{FP}$ value of a particular diode is also determined by the type of material used in its construction. For example, germanium devices will have a typical $V_{FP}$ of 500 millivolts at 25° centigrade.

Manufacturers sometimes define the negative resistance region of a device as a *negative conductance* $-G_d$. The negative conductance is determined by dividing a change in forward current ($\Delta I$) by a corresponding change in forward voltage ($\Delta V$) and is expressed in a unit known as the mho. The negative conductance of a device therefore provides an indication of the slope (rate of change of current with respect to voltage) of the curve within the negative resistance region. In some cases the manufacturer will specify the reciprocal of the negative conductance or the *negative resistance* $-R_d$ of the device. The negative resistance of the diode is therefore equal to $1/-G_d$ and is expressed in ohms. For example, if $\Delta I =$ 5 milliamps and $\Delta V = 5$ volts, then negative conductance is:

$$-G_d = \frac{\Delta I}{\Delta V} = \frac{5mA}{5V} = 0.01 \text{ mho}$$

Knowing negative conductance, you can calculate negative resistance as:

$$-R_d = \frac{1}{-G_d} = \frac{1}{0.01\text{ mho}} = 100\text{ ohms}$$

A tunnel diode may be manufactured to have a negative conductance that is within minimum and maximum limits of .0065 to .01 mho. This same diode would therefore have a negative resistance that varies from 100 to 150 ohms.

## Construction

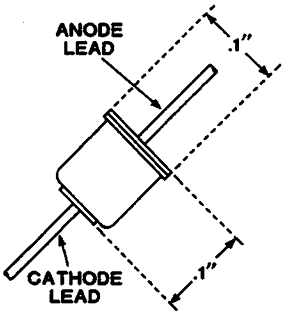


Figure 4-2
Typical tunnel diode package

Tunnel diodes may be constructed from several types of semiconductor materials. Basic semiconductor materials such as germanium and silicon have been used for years but many of the newer devices are constructed from gallium arsenide and gallium antimonide. The tunnel diode PN junctions may be formed by using the grown method, the diffusion method, or the alloyed method. However, the alloyed method is perhaps the most widely used construction technique. Many tunnel diodes resemble ordinary PN junction diodes or zener diodes in appearance but some tunnel diodes are packaged in special cases or containers which make them suitable for various applications. A typical tunnel diode and its approximate dimensions are shown in Figure 4-2. This particular diode is packaged in a metal case and has a peak current rating in the low milliampere range. Since the device is only one tenth of an inch long and one tenth of an inch wide, it is smaller than the head of a match.

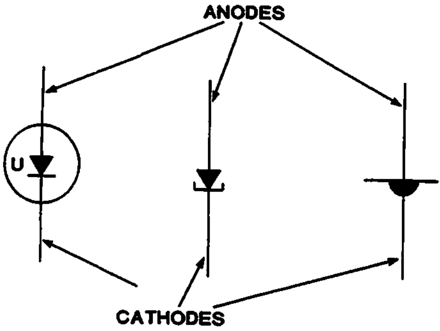


Figure 4-3
Tunnel diode symbols.

Several commonly used tunnel diode symbols are shown in Figure 4-3. Two of the symbols resemble the conventional PN junction diode symbol, but the third symbol is completely different. This symbol consists of a bar, which represents the anode, and a half circle, which represents the cathode. This symbol is also occasionally drawn within a full circle like the diode symbol shown on the left.

## Applications

The tunnel diode is smaller and faster in operation than the transistor. The high switching speed of tunnel diodes, coupled with their simplicity and stability, is suitable for high speed operation. The tunnel diode operates effectively as an amplifier, oscillator, and convertor of high frequency ac signals. A typical tunnel diode oscillator circuit is shown in Figure 4-4. When used in this manner the diode must be biased so that it will operate within its negative resistance region. Notice that the diode is connected in series with an LC resonant circuit. The circuit receives its power from the battery, and the two resistors ($R_1$ and $R_2$) are used to set the diode's operating current and voltage into the negative resistance region.

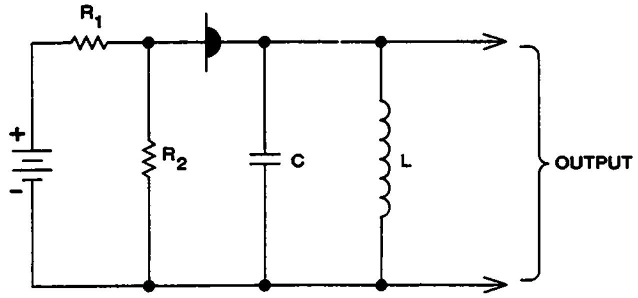


Figure 4-4

Typical tunnel diode oscillator circuit

The resonant LC circuit cannot sustain oscillations when it is used alone. However, when the LC circuit is used with the tunnel diode as shown, continuous oscillations are produced and an ac output voltage can be taken from the resonant circuit. The continuous oscillations result because of the diode's negative resistance. When power is first applied to the circuit, oscillations are produced with the LC circuit. These oscillations produce a voltage across the LC circuit and this voltage alternately causes a shift in the diode's operating point. This in turn causes the diode's resistance to change in a manner which allows the current through the diode to reinforce the circulating current that flow through the resonant LC circuit.

Therefore, the tunnel diode's negative resistance is used to support the oscillations that are produced within the LC circuit and the power losses within the LC circuit are effectively reduced to zero.

The tunnel diode may also be used as an electronic switch. When used in this manner the device is made to change between two-operating states. In one state the device conducts a relatively high forward current at a point just before $I_P$ is reached. In the other state it conducts a relatively low current at a point just beyond the $I_V$ value of the device. Tunnel diodes have been used to implement switching circuits that perform high speed digital logic functions.

When used in either of the applications just described, the tunnel diode requires very little power and it is capable of operating at very high speeds. When used as an oscillator the tunnel diode is capable of operating in the microwave frequency range (above 200 Megahertz) and when used as a switch the device can change states in only a few nanoseconds. Unfortunately, tunnel diodes also have disadvantages which have seriously limited their use in many applications. In general, the tunnel diode's important electrical characteristics vary widely with changes in temperature and their operation is greatly affected by changes in operating voltages. These two factors make it extremely difficult to stabilize the operation of tunnel diode circuits.

## Self Test Review

1. A tunnel diode will conduct high reverse currents when it is reversed-biased.

   a. True
   b. False

2. A tunnel diode's negative resistance region occurs when the device is ______________-biased.

3. Throughout the tunnel diode's negative resistance region, forward current ______________ as forward voltage increases.

4. A tunnel diode's peak current $I_P$ represents the maximum current that the diode can safely handle.

   a. True
   b. False

5. The minimum forward current that flows through a tunnel diode before it functions like a conventional diode is referred to as the diode's ______________ current.

6. Tunnel diodes are usually designed to have a high ______________ current ratio.

7. The negative conductance of a tunnel diode provides an indication of the ______________ of the diode's V-I curve within the negative resistance region.

8. A tunnel diode's projected peak voltage ($V_{FP}$) is the forward voltage that appears across the diode when the diode's forward current rises to a value that is again equal to the diode's ______________ current.

9. A tunnel diode's negative resistance can be determined by taking the reciprocal of its negative conductance.

   a. True
   b. False

10. When used in an oscillator circuit, the tunnel diode is biased to operate within its ______________ ______________ region.

## Answers

1. a. True
2. forward
3. decreases
4. b. False — A tunnel diode's peak current ($I_P$) represents the maximum forward current that flows through the diode before the negative resistance region is reached.
5. valley
6. peak-to-valley
7. slope
8. peak
9. a. True
10. negative resistance

# THE VARACTOR DIODE

When any PN junction diode is reverse-biased, majority carriers are swept away from the diode's junction and a relatively wide depletion area is formed. When the diode is subjected to varying reverse bias voltage, the width or thickness of this depletion region will also vary. When the reverse bias voltage increases in value, the depletion region becomes wider. When the reverse voltage decreases, the depletion region becomes narrower. The depletion region acts like an insulator since it provides an area through which no conduction can take place. It also effectively separates the N and P sections of the diode in the same way that a dielectric separates the two plates of a capacitor. In fact, the entire PN junction diode is basically a small electronic capacitor that changes its capacitance as its depletion region changes in size.

Ordinary PN junction diodes possess only a small amount of internal junction capacitance and in most cases this capacitance is too small to be effectively used. However, special diodes are constructed so that they have an appreciable amount of internal capacitance and are used in much the same way that ordinary capacitors are used in electronic circuits. These special diodes are commonly referred to as *varactor diodes* or variable-capacitance (varicap) diodes.

## Electrical Characteristics

A varactor is usually operated with a reverse bias voltage ($V_R$) that is less than its reverse breakdown voltage rating. As the reverse voltage is increased, the depletion region within the device widens and therefore acts as a wider dielectric between the N and P sections of the device. Since the value of any capacitor varies inversely with the thickness of the dielectric between its plates, the diode's junction capacitance will decrease as the reverse voltage increases. This means that a decrease in reverse bias voltage will cause an increase in the varactor diode's internal junction capacitance. The varactor diode's capacitance therefore varies inversely with the reverse bias voltage applied to the device. However, the capacitance does not vary in a linear manner, and therefore, is not inversely proportional to the reverse bias voltage. The nonlinear change in capacitance that occurs in a typical varactor diode is shown in Figure 4-5. Notice that the capacitance drops rapidly from a maximum value of approximately 40 picofarads as the reverse bias voltage is increased from zero volts. However, the capacitance levels off to a value of approximately 5 picofarads when the reverse bias voltage approaches 60 volts. Although not shown in Figure 4-5, the capacitance of the diode will actually increase above 40 picofarads if the diode is forward biased and will continue to increase as forward bias is increased. However, a point is soon reached where the diode's barrier voltage is overcome (0.6 or 0.7 volts for silicon diodes) and the diode will conduct a high forward current. This forward current makes the diode useless as a capacitor and operation in forward conduction region must therefore be avoided.

Also, should the reverse breakdown voltage of the diode be reached the device would again become useless as a capacitor because of the high reverse current that would occur.

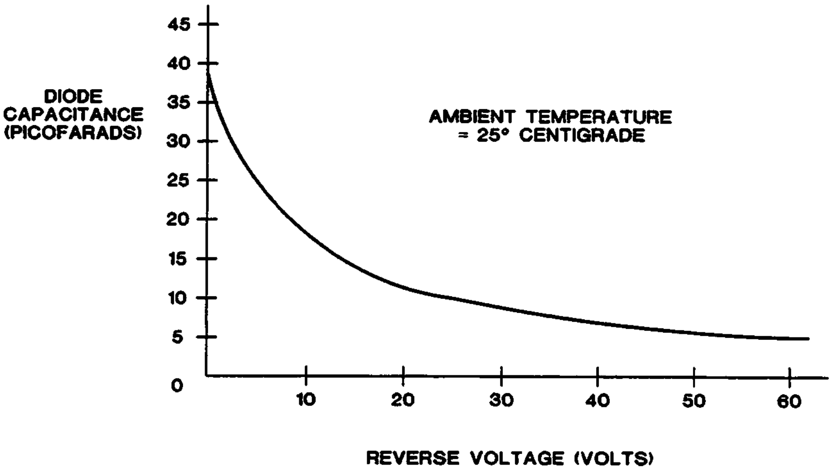


Figure 4-5

Typical capacitance versus voltage characteristics of a varactor diode

A varactor's internal capacitance will increase or decrease slightly when temperature increases or decreases. Therefore, manufacturers of these devices will usually show a range of capacitance values at a particular operating temperature. For example, the curve shown in Figure 4-5 was plotted at an ambient temperature of 25° centigrade. The amount of capacitance change that can be expected for a given change in temperature is usually expressed in a *temperature coefficient of capacitance* $TC_C$. The $TC_C$ is expressed in percent of capacitance change per degree centigrade.

# Operating Efficiency

The relative efficiency of a capacitor is expressed as a ratio of the energy stored by the device to the energy actually used or dissipated by the device. This ratio is referred to as the *quality factor* or Q of the capacitor and it can be determined by dividing the capacitive reactance $X_C$ of the capacitor by the series resistance $R_S$ of the device. However, $X_C$ varies with frequency and the Q will therefore change as frequency changes. Since $X_C = 1/2\ \pi fC$, the Q can be expressed in more specific mathematical terms as

$$Q = \frac{1}{2\pi fCR_S}$$

A varactor diode has internal capacitance and as far as an ac signal is concerned the device also has an internal resistance. This internal resistance is due to the bulk resistivity of the semiconductor material and even though the device is reverse-biased and effectively blocks dc current, this internal resistance is seen by ac signals which pass through the device because of its capacitance. This internal resistance appears as a small resistance in series with the device. This means that varactors have a measurable Q just like ordinary capacitors. In fact, most varactors have very high Q's.

The equation above shows that the Q of any capacitor decreases as frequency, capacitance, or series resistance increases. Therefore, the Q of a varactor must be specified for particular operating conditions if it is to provide a meaningful indication of the diode's efficiency. Since the capacitance of a varactor varies inversely with reverse voltage, the Q of the device will increase as the reverse bias voltage is increased. The maximum Q is therefore obtained just before the breakdown voltage is reached. Often manufacturers will provide graphs which plot Q values as well as capacitance values over a range of reverse bias voltages. In other cases manufacturers will simply indicate the Q of a particular device at a specific frequency (often 50 megahertz) and a specific reverse bias voltage. When comparing Q values of various varactor diodes, always remember that the specific operating conditions must also be considered. If a varactor is subjected to a sufficiently high frequency the Q of the device can be reduced to a value of 1. The frequency at which this point occurs is commonly referred to as the *cutoff frequency* $f_{CO}$ of the device. This cutoff frequency may be determined for various values of C and $R_S$ by using the following equation:

$$f_{CO} = \frac{1}{2\pi CR_S}$$

Manufacturers of varactor diodes will often specify the maximum *leakage current* $I_R$ of each device since this current could have a serious effect on the Q of the device and the manner in which it will operate in various circuits. This leakage current is usually extremely small (in the low microampere range). In order to obtain a high Q, the leakage current must be very low. For this reason most varactor diodes are made from silicon although some units that are designed to operate at microwave frequencies are made from gallium arsenide.

# Construction

Varactor diodes are designed for a number of specific applications. They come in a variety of sizes and shapes and have various power, current, and voltage ratings. These diodes may have capacitance values that range from less than 1 picofarad to one or two thousand picofarads. General purpose, low power diodes are available which closely resemble the ordinary PN junction diodes described in a previous unit. These low power units are often rated at several hundred milliwatts and they are usually packaged in a glass case with axial leads. Other types are available that have higher power ratings (25 to 35 watts) and are stud mounted just like power diodes. These are commonly used at frequencies below 500 Megahertz. At frequencies much above 500 Megahertz the physical layout of the diode package becomes extremely important. Therefore most of the diodes that operate at these higher microwave frequencies are usually contained in packages that are designed to minimize stray or unwanted capacitance and inductance which can cause undesirable effects at these frequencies. Two varactor diodes that are designed to operate at frequencies well above 500 Megahertz are shown in Figure 4-6. These devices are extremely small, as indicated by the approximate dimensions shown, and have power dissipation ratings of 500 milliwatts.

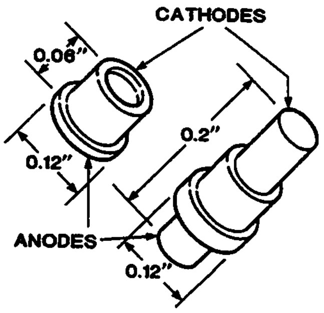


Figure 4-6
Typical high frequency varactor diode package.

It is important to realize that the power dissipation rating of a varactor diode does not have the same meaning as it does in the case of an ordinary PN junction diode. The power dissipated by an ordinary diode is equal to the product of the diodes forward dc current and its forward voltage. However, a reverse-biased varactor diode effectively does not allow a reverse current to flow, but it will pass an ac current. It is this ac current that determines the amount of power dissipated by the device because it flows through the series resistance ($R_S$) of the device and causes heat to be generated. The power dissipated by the device is therefore based on ac calculations rather than dc calculations. Varactor diodes may also be rated according to the amount of ac input power that they can accept or the amount of ac input power that they can produce when used in specific applications.

Several schematic symbols are commonly used to represent the varactor (varicap) diode are shown in Figure 4-7. Notice how all of the symbols incorporate a diode and a capacitor symbol to represent the varactor.

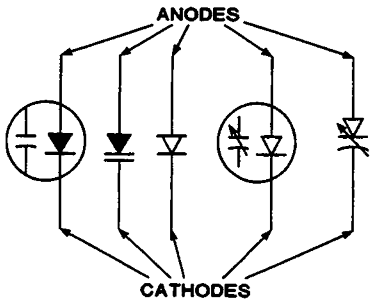


Figure 4-7
Commonly used varactor diode symbols.

## Applications

Varactor diodes are used to replace conventional capacitors in many applications. A varactor diode is often used to vary the frequency of a resonant circuit. In this application the diode may represent the only capacitance in the circuit or represent only part of the capacitance of the circuit. The frequency of the resonant circuit is controlled by simply varying the diode's reverse bias voltage. This means that the diode must be connected so that it is properly biased within the circuit. A basic circuit arrangement is shown in Figure 4-8. Capacitor $C_1$ is quite large and therefore has a very low reactance at the resonant frequency of the parallel resonant circuit which consists of the varactor diode and the inductor L.

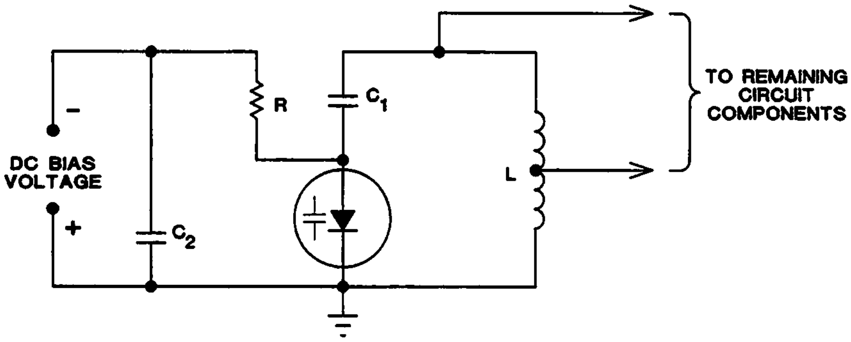


Figure 4-8 Controlling frequency with a varactor diode

Capacitor $C_1$ is used only to prevent the dc bias from flowing through the inductor. The dc voltage, resistor R and capacitor $C_2$ are used to reverse bias the diode as shown. The capacitance of the varactor diode changes as the bias voltage is varied and therefore changes the frequency of the resonant circuit. When used in this manner the varactor diode becomes a tuning component and is often called a tuning varactor or voltage variable capacitor.

Varactor diodes may also be used in many other applications. For example, they find extensive use in various types of high frequency amplifiers and in devices known as frequency multipliers. They are also used in the automatic frequency control (AFC) circuits found in many FM radios and they are used in a variety of circuits that are found in both AM receivers and transmitters.

# EXPERIMENT 4

# Varactor Diodes

*OBJECTIVE:* *1. To demonstrate the characteristics of a varactor diode.*

*2. Investigate a varactor diode's operation as a capactitor.*

## Introduction

This experiment demonstrates the operational characteristics of a typical voltage-variable capacitance diode. In the experiment, you will plot the relationship between the reverse bias voltage, the capacitance of a varactor diode and the capacitive reactance of the circuit.

## Material Required

Heathkit Analog Trainer
DC voltmeter (VOM)
AC ammeter (may be part of VOM)
1 — 2.2 kilohm resistor (red, red, red, gold)
1 — Varactor diode MV209 (56-674)
1 — 470 ohm resistor (yellow, violet, brown, gold)

## Procedure

1. Turn on the Trainer. Adjust the + voltage control for 15 volts. Turn the Trainer off.

   **Note:** In the next few steps you will adjust the 1 kilohm potentiometer until you attain a specified voltage level, then you will record the current. There are two ways in which this sequence can be accomplished. If your voltmeter and ammeter are separate pieces of equipment, you can connect both meters to the locations shown and keep them connected as you adjust the potentiometer, and make your measurement. If your voltmeter and ammeter are the same piece of equipment, you will have to switch between the two modes of operation. Remember to turn the Trainer off before inserting or removing the ammeter in the circuit.

2. Construct the circuit shown in Figure 4-9. If your voltmeter and ammeter are the same piece of equipment, do not install the ammeter in the circuit at this time. If you are using an ammeter, it should be set to measure 20 mA of ac current.

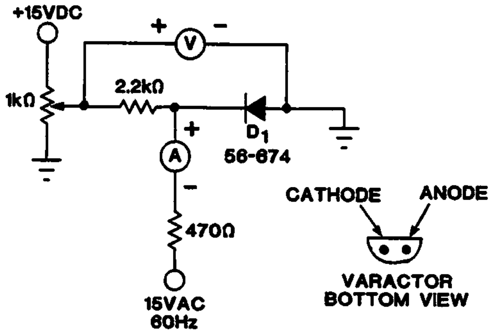


Figure 4-9
Varactor diode test circuit.

3. Turn the Trainer on and adjust the 1 kilohm potentiometer for a 0 Vdc measurement. The potentiometer is used to control the reverse bias voltage ($V_{RB}$).

4. If necessary, insert the ac ammeter in the circuit and set it to measure 20 ma. The ammeter should indicate approximately 10.5 mA. Record the actual current in Figure 4-10, next to the 0 Vdc $V_{RB}$ entry.

| $V_{RB}$ | I (mA) | $X_C$ | Capacitance |
|---|---|---|---|
| 0 Vdc | | | |
| 2 Vdc | | | |
| 4 Vdc | | | |
| 6 Vdc | | | |
| 8 Vdc | | | |
| 10 Vdc | | | |

Figure 4-10
Chart for recording the varactor circuit characteristics.

5. Using the 1 kilohm potentiometer, adjust the reverse bias voltage ($V_{RB}$) to 2 Vdc.

6. Record the current reading in Figure 4-10.

7. Continue to adjust the voltage and measure the current until you have completed the current column in Figure 4-10.

8. After you complete the measurements, turn off the Trainer and disconnect the circuit.

9. Calculate the circuit reactance ($X_C$) for each dc voltage setting using the equation $Vac/I = X_C$. Record this data in Figure 4-10.

10. Transfer the data from Figure 4-10 into the graph in Figure 4-11.

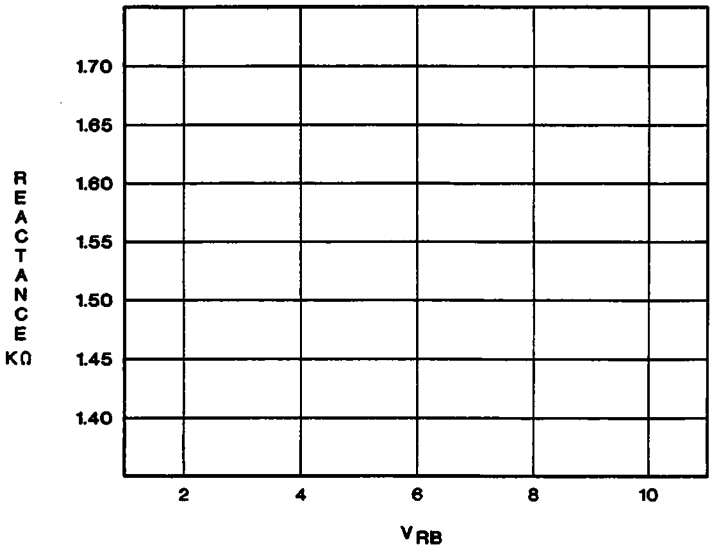


Figure 4-11
Circuit Reactance (kilohms) verses reverse bias voltage.

11. Calculate the varactor capacitance for each dc voltage setting using the equation $1/2\pi f X_C = C$. Record the capacitance in Figure 4-10. Transfer the data from the chart in Figure 4-10 to the graph in Figure 4-12.

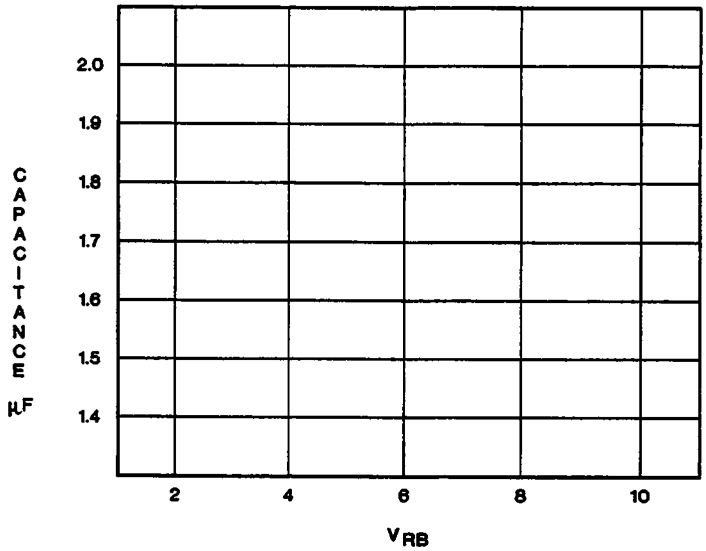


Figure 4-12
Varactor Capacitance (μF) verses reverse bias voltage.

## Discussion

You discovered that as you increase the dc reverse bias voltage, the current decreases. Decreasing the current caused the circuit reactance to increase. In turn, as the reactance increases, varactor capacitance decreases. These facts become more evident when you view the graphs in Figures 4-11 and 4-12. With this in mind, it is easy to see how a dc bias voltage can be used to vary the capacitance of a varactor. It is for this reason varactor diodes are often used in a tuned circuit. The 15 Vac 60 Hz signal is used to drive the circuit.

# HIGH FREQUENCY DIODES

There are a variety of special diode types designed for use in high frequency applications. These include both switching (on-off) types and radio frequency signal generating types. All of these diodes operate in the VHF, and UHF and microwave regions. Ins this section you will examine some of these diodes. The most important types include the PIN, impatt, Schottky and Gunn effect diodes.

## PIN Diodes

All of the diodes previously described utilize a PN junction which is formed by doping a single semiconductor material with different impurities. Each of these diodes contain just two oppositely doped sections which meet to form a single junction. However, another special type of diode is constructed in a slightly different manner. This diode contains an undoped or intrinsic (pure semiconductor) region which is sandwiched between heavily doped N and P sections as shown in Figure 4-13. Since this diode contains a P-type layer, an intrinsic or I layer, and an N-type layer, it is commonly referred to as a PIN diode.

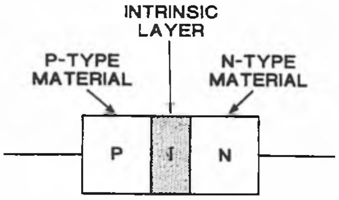


Figure 4-13

PIN diode construction

PIN diodes are capable of changing from one operating state to another at an extremely fast rate and are therefore used in high-speed switching and pulse generation applications. When used as a switch, the PIN diode is turned on or off by subjecting it to forward or reverse bias voltages. When a PIN diode is forward-biased, it's forward resistance will change considerably as its forward voltage is varied. Although most other diodes exhibit a variable resistance characteristic, they do not have the wide resistance range of the PIN diode. Furthermore, the internal resistance of the PIN diode changes linearly with forward voltage and the device responds well to low bias voltages and currents. The resistance of a typical PIN diode may change from 10,000 ohms to 1 ohm when its forward current changes from 1 microampere to 100 milliamperes. This characteristic makes the device suitable for use in certain types of limiter circuits or they may simply be used as attenuators (current-controlled resistors). In general, PIN diodes find their greatest applications at the higher (microwave) frequencies. Both low power and high power devices are available. These devices are often packaged like the varactor diodes shown in Figure 4-6, but other types are available. For example, the devices shown in Figure 4-14 are actually PIN diodes. These diodes are packaged in hermetically sealed metal cases which serve as heat sinks. The case also provides an electrical connection to one end of the diode within the case (the cathode or anode end) and the protruding metal lead provides the other diode connection. For the actual electrical characteristics refer to a specific data sheet.

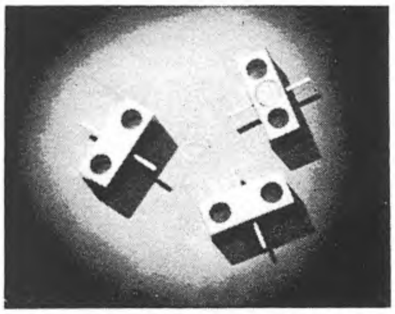


Figure 4-14
Hermetically sealed PIN diode packages.
(Courtesy of Hewlett Packard).

A special type of PIN diode, which exhibits an abrupt change in internal resistance, is also used. This device is often referred to as a *step recovery diode* or SRD. The SRD conducts heavily when forward-biased until all of the charge carriers within its I region are depleted. Then its internal resistance jumps to a high value causing the forward current through the device to immediately drop to a low level. SRDs are used in various types of pulse shaping circuits and they are also used to generate narrow or sharp voltage pulses.

## Schottky Diodes

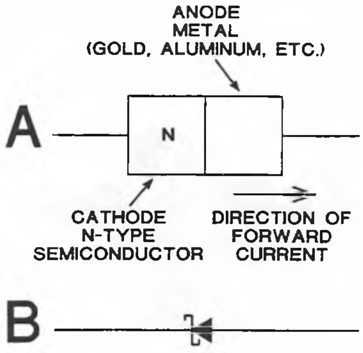


Figure 4-15
A Schottky diode and its schematic symbol.

The Schottky diode or Schottky barrier is formed by placing an N-type semiconductor material (usually silicon) in contact with a metal such as gold, silver or aluminum to form a metal-to-semiconductor junction. This diode operates in a manner similar to ordinary PN junction diodes but there are several important differences. The barrier voltage developed within the device is approximately one half as great as the barrier voltage within an ordinary silicon diode. This means that the forward voltage drop across the diode is approximately 0.3 volts instead of 0.6 or 0.7 volts. Also, the Schottky diode operates with majority carriers (electrons); virtually no minority carriers are involved. This means that the reverse or leakage current through the device is extremely small. Figure 4-15 shows the basic Schottky diode construction and the symbol used to represent it.

Schottky diodes are sometimes called hot carrier diodes (HCD). The name hot carrier diode is used because, in a forward-biased HCD, the electrons possess a high level of kinetic energy as they move from the N-type material across the junction to the metallic anode.

The Schottky diode is able to change operating states (turn on and off) much faster than ordinary PN junction diodes, and it is used extensively to process high frequency ac signals. This device finds extensive use in microwave electronic mixers (circuits which combine ac signals), detectors (circuits which use rectification as a means of extracting information from ac signals), and high speed digital logic circuits. Figure 4-16 shows some typical diodes for use in microwave circuits.

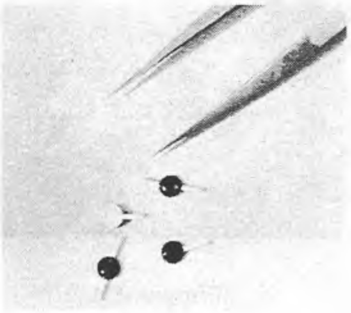


Figure 4-16
Typical Schottky barrier diode packages.
(Courtesy of Hewlett Packard).

# Gunn-Effect Devices

Gunn-effect diodes are capable of producing oscillations when they are used in conjunction with a resonant circuit and a dc operating voltage. These diodes are usually made of N-type gallium arsenide (GaAs) semiconductor crystals and do not have a PN junction like an ordinary diode. However, these devices display a negative resistance characteristic within their bulk semiconductor material.

Application of a voltage across a slice of N-type GaAs produces an electric field that causes current to flow toward the positive end. It is a characteristic of N-type GaAs that electrons move through the material at high velocity. As the voltage is increased toward a *threshold* voltage, current and velocity of the electrons also increase. If the voltage is increased beyond threshold, the device begins to exhibit negative resistance as a result of a decrease in electron velocity and current. This is because electrons gain so much energy that they actually change to a higher, ordinarily empty, energy band. This phenomenon is the reason that the term *transferred electron effect* is sometimes used to describe the Gunn effect — electrons have been transferred from the conduction band to a higher energy band in which they move more slowly.

When electrons have been transferred to this higher energy band, they enter the cathode more rapidly than they leave, resulting in the gathering of electrons at the cathode end in an accumulation called a *domain.* The domain produces an electric field, concentrated at the cathode end, falling off to a level below threshold at the anode end. As shown in Figure 4-17A through F, the domain drifts across the device to the anode and leaves a pulse, and the cycle repeats (the X axis in Figure 4-17 represents the thickness of the Gunn diode). As a consequence, the output from the Gunn diode is a series of current pulses whose frequency is determined by the width of the drift zone. These pulses can be used to generate microwave signals when they are applied to a resonant cavity (a resonant cavity is essentially a tuned LC circuit that is capable of oscillating at microwave frequencies).

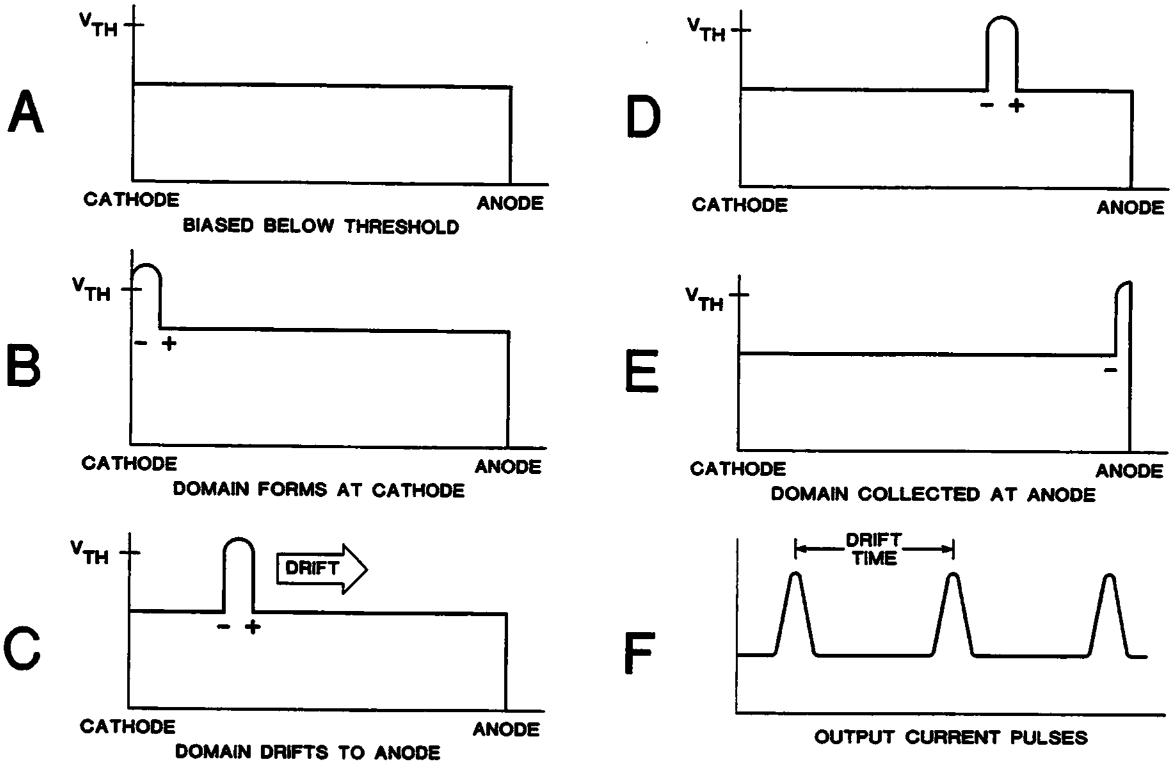


Figure 4-17
Gunn Diode domain drift.

# IMPATT Diodes

The construction of an IMPATT (Impact Avalanche Transit Time) diode is similar to that of a zener diode and like a zener diode, an IMPATT diode can withstand avalanche current. Figure 4-18 shows a PN junction that is reverse biased with a voltage sufficient to cause the diode to be in avalanche breakdown. During avalanche breakdown, a small increase in applied voltage causes a large increase in current. If an external resonant circuit applies a high frequency RF voltage to the reverse biased junction, the RF voltage alternately aids and opposes the reverse bias. As a consequence, avalanche current is alternately aided and opposed. However, avalanche current changes are not instantaneous because, as indicated in the Figure 4-18, a large number of electrons move from the avalanche zone into the drift zone. As a result, a phase difference is introduced between the applied voltage changes and avalanche current changes. If the RF current and RF voltage are out of phase by more than 90 degrees, the junction is exhibiting negative resistance to the RF signal. This is because, as the RF voltage increases, the RF current decreases and vice versa.

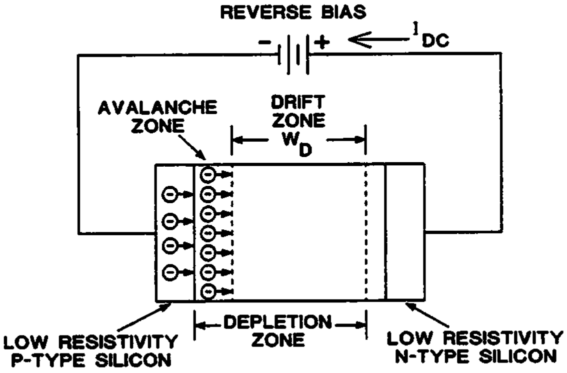


Figure 4-18
IMPATT Diode input and output phase relationships

In a practical application, IMPATT diodes are reverse biased just above the avalanche breakdown point, $V_B$ In Figure 4-19A. This allows the positive transistions of the applied rf voltage to reverse bias the diode into avalanche. As shown in Figure 4-19B, avalanche current then builds up and peaks when the applied voltage returns to zero volts at 180 degrees through its cycle. As the applied voltage goes through its negative half cycle, avalanche decreases, producing a sharply peaked current pulse as shown in Figure 4-19C. This pulse then drifts toward the positively charged side of the junction at an almost constant speed. The time required for this pulse to pass through the drift zone is determined by the width of the drift zone. Therefore, the frequency of the output pulses is determined by controlling the width of the drift zone at the time of manufacture.

As the current pulse drifts through the diode, it induces a positive square waveform of current into the external resonant circuit, as indicated in Figure 4-19D. This square wave is present during the negative half cycle of the applied RF signal. Therefore, it is 180 degrees out of phase with the input signal. Thus, the diode generates RF energy as a result of negative resistance.

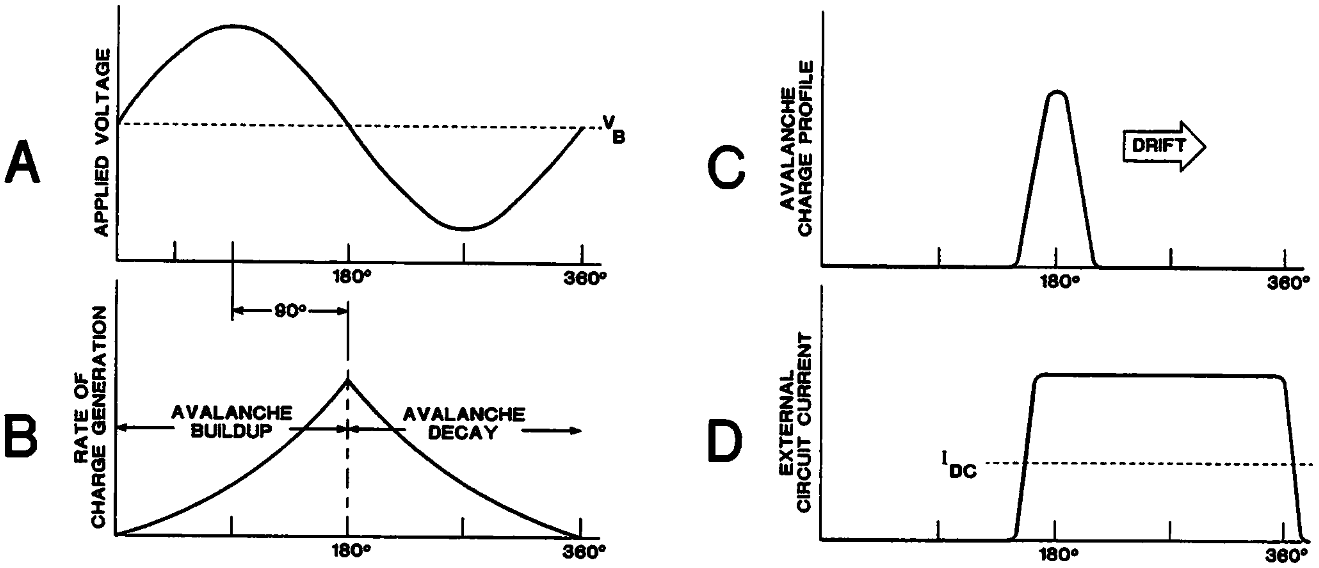


Figure 4-19
IMPATT Diode operating characteristics.

IMPATT diodes are primarily used to produce microwave radio frequencies above 3000 MHz and are capable of producing higher power levels than Gunn diodes. Silicon, gallium arsenide and, occasionally, germanium are used in the manuafacture of IMPATT diodes. A disadvantage of IMPATT diodes is that they produce higher noise levels than Gunn diodes. However, noise can be minimized by using GaAs IMPATT diodes and proper circuit design. Another disadvantage of IMPATT diodes is that they require relatively high operating voltage.

# Self Test Review

23. A diode which contains a P-type material, an intrinsic layer, and an N-type material is referred to as a ______________ diode.

24. The diode packages shown in Figure 4-6 are typical of those devices that are designed to operate at ______________ frequencies.

25. IMPATT diodes provide a means of generating __________ __________ __________.

26. The Schottky diode utilizes a metal-to-semiconductor junction.

 a. True
 b. False

27. The barrier voltage within an HCD is approximately equal to __________ volts.

28. Gunn-effect diodes and IMPATT diodes are capable of generating RF power without the use of external components or voltages.

 a. True
 b. False

29. A special type of PIN diode which exhibits an abrupt change in internal resistance is called a ______________ ______________ diode.

30. IMPATT diodes operate in the reverse biased condition.

 a. True
 b. False

31. To forward bias a Schottky diode, the N-type semiconductor element is made ______________ with respect to the metallic element.

## Answers

23. PIN
24. extremely high or microwave
25. extremely high frequencies.
26. a. True
27. 0.3
28. b. False — both of these diodes must be used with a tuned resonant circuit and an external dc voltage.
29. step recovery
30. a. True
31. negative

# UNIT SUMMARY

Modification of the construction and/or doping of the basic PN junction makes it possible to produce semiconductor diodes with characteristics that make them useful for special applications.

The PN junction of a tunnel diode is heavily doped. This results in a high internal barrier voltage and a very narrow depletion region, as well as a low reverse breakdown voltage. One of the most significant characteristics of tunnel diodes is the negative resistance region of their operating curves. This negative resistance characteristic makes tunnel diodes useful as part of microwave oscillator circuits.

A certain amount of capacitance is present in all PN junction diodes when they are reverse biased. This capacitance results from the depletion region acting as an insulator between the two conductive regions of the diode. Varactor diodes are specially constructed to make use of this capacitance. Varying the reverse bias voltage on a varactor diode causes the capacitance of the diode to change. Consequently, varactor diodes are also called voltage variable capacitors. Varactor diodes are frequently used to control the frequency of resonant circuits.

In a PIN diode, the N-type region and the P-type region are separated by a layer of intrinsic (I) material. A reverse-biased PIN diode does not conduct and acts like a capacitor. When forward bias is applied, the diode acts as a low series resistance. One of the most frequent applications of PIN diodes is as an electronic switch.

Schottky diodes consist of an N-type material in contact with a metal. This construction results in an internal barrier voltage that is approximately one half as great as the barrier voltage in an ordinary PN junction. These devices conduct primarily by means of electrons, with practically no hole current. As a consequence, leakage current through a reverse biased Schottky diode is extremely small. Schottky diodes are used in microwave mixer and detector circuits and in high speed logic circuits. Schottky diodes are also known as Schottky-barrier diodes and Hot Carrier Diodes (HCD).

Gunn diodes have a negative resistance characteristic similar to that of a tunnel diode even though both devices operate on a different principle. The operation of a Gunn diode does not depend on a PN junction or contact characteristics. N-type gallium arsenide is the most commonly used material used in the fabrication of Gunn diodes. However, several other materials that display the Gunn effect are also used. Accumulations of electrons called domains are produced by an electric field within the device. As a result, Gunn diodes produce a pulsed current output. The primary application of Gunn diodes is in microwave oscillator circuits.

IMPATT diodes, like Gunn diodes, produce a pulsed current output. For this reason, they are also used as part of microwave oscillator circuits. However, IMPATT diodes are capable of producing higher output power than Gunn Diodes. IMPATT diodes make use of a PN junction that is reverse biased to cause the diode to operate as an avalanche breakdown device. This enables the negative resistance of the device to introduce a 180 degree phase shift between the input signal and the output current pulses. IMPATT diodes are usually used as part of a microwave oscillator circuit. They are capable of producing higher rf output power than other devices. However, they also produce more noise. IMPATT diodes can be made of gallium arsenide, silicon, and less frequently, germanium.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear in the Examination Answer Sheet which follows.

1. The tunnel diode has a PN junction that is

   A. separated by an intrinsic layer.
   B. separated by a pure metal.
   C. lightly doped.
   D. heavily doped.

2. A tunnel diode is operating within its negative resistance region when its forward current

   A. decrease as forward voltage increases.
   B. remains constant as forward voltage increases.
   C. increases as forward voltage increases.
   D. decreases as forward voltage decreases.

3. If a tunnel diode is to have a wide operating range within its negative resistance region it must have

   A. high valley current.
   B. low peak current.
   C. low peak-to-valley current ratio.
   D. high peak-to-valley current ratio.

4. A varactor diode's internal capacitance increases as the diode's

   A. leakage current increases.
   B. depletion region widens.
   C. reverse bias voltage decreases.
   D. reverse bias voltage increases.

5. Varactor diodes have a measurable Q because they have internal

   A. capacitance only.
   B. capacitance and resistance.
   C. capacitance and inductance.
   D. capacitance.

6. The amount of power dissipated by a reverse-biased varactor diode is determined by the amount of

   A. dc current flowing through the device.
   B. dc voltage across the device.
   C. ac and dc current flowing through the device.
   D. ac current flowing through the device.

7. The varactor diode becomes useless as a capacitor when it's

   A. subjected to any reverse-bias voltage.
   B. subjected to a reverse-bias voltage that is just below its breakdown rating.
   C. operated in its forward conduction region.
   D. subjected to very low reverse-bias voltages.

8. The PIN diode contains an N-section, a P-section and

   A. a metal layer.
   B. an intrinsic layer.
   C. a region that is heavily doped.
   D. a region that is lightly doped.

9. The IMPATT diode is used to

   A. provide a regulated output voltage.
   B. generate high power, low frequency signals.
   C. generate extremely high frequencies.
   D. provide rectification of low frequency signals.

10. The hot carrier diode utilizes a

    A. semiconductor-to-metal junction.
    B. standard PN junction.
    C. metal-to-metal junction.
    D. PIN junction.

11. Which type of diode does *not* normally operate in the reverse biased position?

    A. IMPATT
    B. tunnel
    C. zener
    D. varactor

12. Which type of diode is *not* normally used to generate high frequency or microwave signals?

    A. IMPATT
    B. tunnel
    C. zener
    D. Gunn

13. Operation of Gunn diodes depends on

    A. the transferred electron effect.
    B. PN junction characteristics.
    C. avalanche breakdown.
    D. minority carriers.

14. The frequency of Gunn diode output pulses is determined by

    A. the threshold voltage.
    B. domains.
    C. drift zone width.
    D. the external circuit.

15. Output pulse width from an IMPATT diode is determined by the

    A. avalanche breakdown voltage.
    B. applied signal frequency.
    C. external circuit.
    D. drift zone width.

# EXAMINATION ANSWERS

1. D — The tunnel diode's heavily doped PN junction helps to give the device its unique electrical characteristics.

2. A — The negative resistance region is between the peak and valley points where forward current decreases as forward voltage increases. The diode's V-I curve has a negative slope throughout this region.

3. D — A high peak-to-valley ration means wide range of currents between $I_P$ and $I_V$.

4. C — The varactor diode's internal capacitance varies inversely (but not inversely proportional) with the diodes reverse bias voltage.

5. B — Varactor diodes have internal capacitance and resistance and therefore function much like an ordinary capacitor. The Q of the diode is equal to the ratio of the diode's capacitive reactance to the diode's resistance.

6. D — Only ac current can flow through a reverse-biased varactor diode and cause the device to dissipate power.

7. C — When the device operates in its forward conduction region, a high forward current flows through the device making it useless as a capacitor.

8. B — The intrinsic or I layer is sandwiched between the P and N sections.

9. C — The IMPATT diode requires only a minimum of additional components to generate frequencies in the microwave region.

10. A — The hot carrier diode is formed by placing an N-type semiconductor in contact with a metal such as gold or silver.

11. B — The tunnel diode is not normally operated in a reverse bias condition but the zener, IMPATT and varactor must be reverse biased for proper operation.

12. C — The zener is not used to generate high frequency signals like the tunnel, IMPATT and Gunn diodes.

13. A — The transferred electron effect is an inherent characteristic of the materials used in the construction of Gunn diodes.

14. C — Gunn diode output frequency is determined by the width of the drift zone.

15. D — Drift zone width controls the output pulse width from an IMPATT diode.

# *Unit 5*

# BIPOLAR TRANSISTOR OPERATION

# CONTENTS

Introduction .......... Page 5-3

Unit Objectives .......... Page 5-3

Unit Activity Guide .......... Page 5-4

PNP and NPN Configurations .......... Page 5-5

Transistor Construction .......... Page 5-6

  Construction Techniques .......... Page 5-6

  Packaging .......... Page 5-8

Basic Transistor Action .......... Page 5-11

  Biasing NPN Transistors .......... Page 5-11

  Biasing PNP Transistors .......... Page 5-14

Transistor Amplification .......... Page 5-16

Transistor Circuit Arrangements .......... Page 5-20

  Common Base Circuits .......... Page 5-20

  Common Emitter Circuits .......... Page 5-21

  Common Collector Circuits .......... Page 5-23

Testing Bipolar Transistors .......... Page 5-27

  Test Procedures .......... Page 5-27

  Testing NPN Transistors .......... Page 5-28

  Testing PNP Transistors .......... Page 5-28

  Identifying a Transistor .......... Page 5-29

Experiment 5 .......... Page 5-32

Experiment 6 .......... Page 5-37

Unit Summary .......... Page 5-42

Unit Examination .......... Page 5-45

Examination Answers .......... Page 5-47

# INTRODUCTION

This unit on bipolar transistor operation, is designed to familiarize you with the bipolar transistor and its basic principles of operation. You will examine some of the basic construction techniques used to form these important components, and learn how the transistors are used to amplify electronic signals. You will also learn a simple method of testing transistors to determine if they are defective. Finally, you will see how a transistor can be used as a switch.

A transistor is a three element electronic component used to control electron flow. The amount of current flowing through a transistor can be controlled by varying the voltages applied to the three elements. By controlling the current, useful applications such as amplification, oscillation, and switching can be achieved.

Examine your unit objectives closely. Then, follow the instruction provided in your unit activity guide to be sure that you perform the necessary steps as you proceed through this unit.

# UNIT OBJECTIVES

When you have completed this unit on bipolar transistors you should be able to:

1. Describe the construction of the two basic types of bipolar transistors.
2. State the principle behind bipolar transistor operation.
3. Show how a bipolar transistor should be biased for normal operation.
4. Define the relationship between emitter current, base current, and collector current in a bipolar transistor.
5. Identify the three basic transistor amplifier circuit arrangements.
6. Test bipolar transistors with an ohmmeter to determine if the devices have internal defects.
7. Construct a common-emitter transistor circuit.

# UNIT ACTIVITY GUIDE

**Completion Time**

☐ Read "PNP and NPN Configurations" and "Transistor Construction." __________

☐ Answer Self Test Review Questions 1-8. __________

☐ Read "Basic Transistor Action." __________

☐ Answer Self Test Review Questions 9-14. __________

☐ Read "Transistor Amplification." __________

☐ Answer Self Test Review Questions 15-19. __________

☐ Read "Transistor Circuit Arrangements." __________

☐ Answer Self Test Review Questions 20-29. __________

☐ Read "Testing Bipolar Transistors." __________

☐ Answer Self Test Review Questions 30-32. __________

☐ Perform Experiment 5 on "Testing Bipolar Transistors." __________

☐ Perform Experiment 6 on "Transistor Characteristics." __________

☐ Read the Unit Summary. __________

☐ Complete the Unit Examination. __________

☐ Check the Examination Answers. __________

# PNP AND NPN CONFIGURATIONS

You are now ready to examine a solid-state component that is related to the PN junction diode that was previously described. This device is called a *bipolar transistor* although it is sometimes referred to as a *junction transistor* or simply a *transistor*. Throughout this unit the words transistor and bipolar transistor have the same meaning and are used interchangeably. Later in this course you will examine another basic type of transistor known as a *unipolar transistor*. The terms *bipolar* and *unipolar* relate to the manner in which transistors operate and transistors in general are classified either as bipolar or unipolar devices.

A bipolar transistor is constructed from germanium and silicon semiconductor materials just like a PN junction diode, however, this device utilizes three alternately doped semiconductor regions instead of two as in the case of the diode. These three semiconductor regions may be arranged in two different ways. One arrangement is shown in Figure 5-1A. Notice that an N-type semiconductor layer is sandwiched between two P-type semiconductor layers to form what is commonly referred to as a PNP type of transistor. The middle or N region is called the base of the transistor and the outer or P regions are called the emitter and collector of the transistor. Normally, the base region is much thinner than the emitter and collector regions and is also lightly doped in relation to the other two. The appropriate leads must also be attached as shown to provide electrical connections to the three regions.

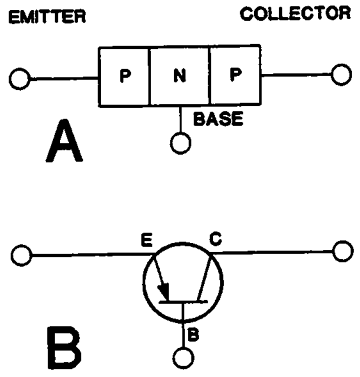


**Figure 5-1**
A basic PNP transistor and its symbol.

The PNP transistor is usually represented in circuit diagrams or schematics by the symbol shown in Figure 5-1B. The emitter, base, and collector of the transistor are identified by the letters E, B, and C respectively; however, these letters may not always appear with the symbol.

The second method of arranging the semiconductor layers in a bipolar transistor is shown in Figure 5-2A. Notice that this time a layer of P-type material is sandwiched between two layers of N-type material to form what is commonly known as an NPN transistor. Like the PNP transistor previously described, this device utilizes a narrow middle region which is referred to as the base. The outer layers are again referred to as the emitter and collector regions.

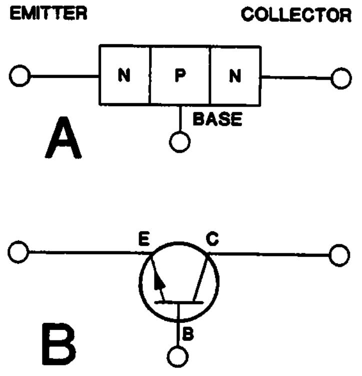


**Figure 5-2**
A basic NPN transistor and its symbol.

The NPN transistor is usually represented by the symbol shown in Figure 5-2B. The respective emitter, base, and collector leads may or may not be identified by the additional letters (E, B, and C) as shown. Notice that the only difference between the PNP and NPN transistor symbols is the direction of the arrow which represents the emitter of the device. The arrow points toward the base in the PNP symbol but away from the base in the NPN symbol.

Because PNP and NPN transistors have three alternately doped semiconductor regions, these devices have two junctions of P and N type materials. In other words each transistor has a PN junction between its emitter and base regions and a PN junction between its base and collector regions. Therefore, bipolar transistors have three layers and two junctions as opposed to PN junction diodes which have two layers and one junction.

# TRANSISTOR CONSTRUCTION

Figures 5-1A and 5-2A show the basic PNP and NPN structures, but still do not show exactly how these devices are constructed. The three semiconductor layers are not just simply pressed together as suggested by these simple illustrations. Instead, the three layers are often formed by using construction techniques that are similar to those used to construct ordinary PN junction diodes. A complete analysis of all of the various construction techniques would be impossible at this time since many different processes have been devised and these processes are constantly being improved. Therefore, we will only briefly review a few basic techniques from which many of the newer techniques have been developed.

## Construction Techniques

In general, bipolar transistors are constructed by using the alloyed method or the diffusion method or by using variations of these basic processes. Figure 5-3A shows a transistor that has been formed by the alloyed method. The three regions are formed by placing indium pellets on each side of a thin N-type semiconductor crystal. The entire unit is then heated until the pellets melt and diffuse into the N-type material. After the heat is removed, the indium (a trivalent material) and the N-type material recrystallize to form a PNP transistor as shown.

Figure 5-3B shows a transistor that has been formed by the diffusion process. This transistor is formed by subjecting an N-type semiconductor crystal to a trivalent impurity at an elevated temperature. The trivalent impurity is allowed to penetrate or diffuse into the crystal to form a P-type region. Then the process is repeated with a pentavalent impurity to form a small N-type region within the P-type region previously formed. This results in an NPN transistor as shown; however, PNP transistors can be formed using the same basic process.

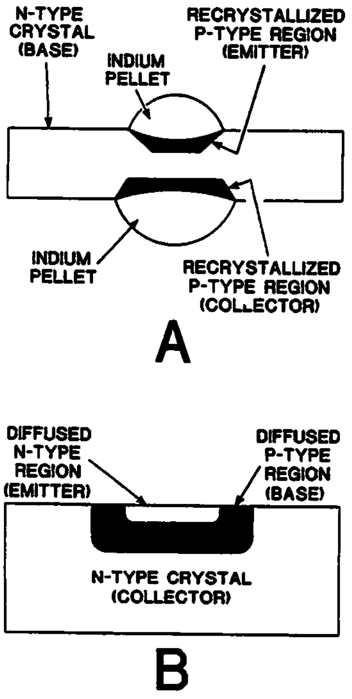

N-TYPE CRYSTAL (BASE)
RECRYSTALLIZED P-TYPE REGION (EMITTER)
INDIUM PELLET
INDIUM PELLET
RECRYSTALLIZED P-TYPE REGION (COLLECTOR)
A
DIFFUSED N-TYPE REGION (EMITTER)
DIFFUSED P-TYPE REGION (BASE)
N-TYPE CRYSTAL (COLLECTOR)
B

**Figure 5-3**
Cross-sections of typical Bipolar transistors.

Both of the construction techniques just described are similar to the techniques used to construct PN junction diodes. Each technique offers certain advantages and disadvantages and both techniques may be combined in various ways to produce transistors with certain desirable characteristics. Also the physical size and shape of the layers is important in determining how the transistor will operate. In general, the base is usually smaller than the collector region. The reasons for this relationship are explained later in this unit.

In addition to the two basic construction methods just described, more refined methods are often used to construct transistors. For example, some transistors are formed through an *epitaxial growth* process. This type of transistor is formed by growing a crystalline film on top of a basic semiconductor crystal. This film has the same crystalline structure as the original material and can contain either type of impurity. This process makes it possible to grow additional P-type or N-type layers to obtain the necessary PNP or NPN arrangement.

Transistors may also be constructed by using the epitaxial growth process along with the diffusion process previously mentioned. The general shape of a transistor is also an important factor. The transistor shown in Figure 5-3B is formed by selectively diffusing the base and emitter regions into the collector region so that the device has a flat top or in other words all of the regions lie in a single plane. When this type of construction is used, the device is referred to as a *planar* transistor. However, in many cases a trench or moat is etched around the base and emitter regions of the transistor which leaves the base and emitter standing on a sort of plateau or mesa. This technique is used to expose the base or emitter so that electrical connections can be easily made to these regions. This type of construction is also used to define the areas where the regions meet. Figure 5-4 shows a transistor that utilizes this *mesa* type of construction as well as epitaxially grown and diffused regions. The construction starts by epitaxially growing a lightly doped P-type layer (the base) on top of an N-type semiconductor crystal (the collector). Then the emitter is formed by diffusing an N-type region into the epitaxial layer. Finally, the area around the emitter is etched away to form a mesa structure.

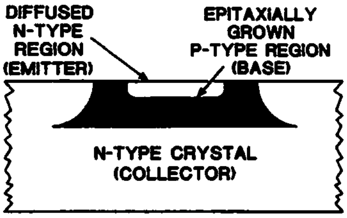

DIFFUSED N-TYPE REGION (EMITTER)
EPITAXIALLY GROWN P-TYPE REGION (BASE)
N-TYPE CRYSTAL (COLLECTOR)

**Figure 5-4**
Cross-section of an Expitaxial-Base transistor which utilizes a mesa structure.

It is important to realize that a large number of transistors are simultaneously formed with any given process. Generally, the transistors are formed on a very thin semiconductor wafer or disc which may be as large as 1½ to 2 inches in diameter. After all of the regions are simultaneously and selectively formed, the wafer is then scribed and broken into many hundreds or thousands of transistors.

## Packaging

Once a transistor is constructed, it must be installed in a suitable container or package which will protect the device and provide a means of making electrical connections to its emitter, base, and collector regions. The package may also serve as a heat sink which will drain heat away from the transistor and prevent it from being damaged. Many different package or case styles are available to cover a wide range of possible applications. Several commonly used transistor packages are shown in Figure 5-5. Packages A through D are designed for low power applications where either small electrical signals or currents are to be amplified or controlled. The transistors are actually located inside of these packages and their emitter, base and collector regions are electrically connected to the leads which protrude from the bottom of each package. Packages A through C are made of metal and the transistors are hermetically sealed within these units to protect them from humidity, dust, and other materials which might contaminate them. The transistor's collector region is often connected directly to the metal case; therefore, one of the leads may serve as the collector lead as well as a case connection.

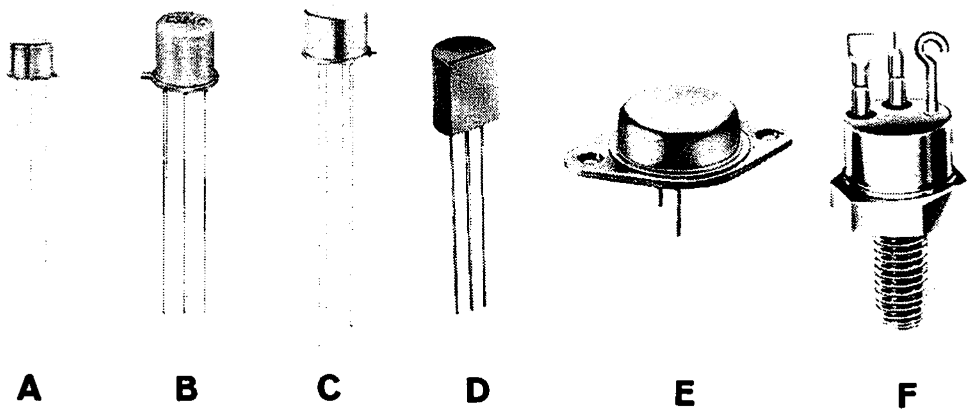


**Figure 5-5**
Typical low power and
high power transistor packages
(Courtesy of Unitrode Corporation)

When it is necessary to have the transistor electrically insulated from the metal case, but still have an electrical connection to the case, package C may be used. With this arrangement, a fourth lead is provided which serves as a case connection.

Package D is made of plastic but similar epoxy types are also available. The transistor is encapsulated within the plastic housing and the three protruding leads connect to the transistor's emitter, base, and collector regions. In general, plastic or epoxy transistors are not as expensive as the metal types, however the plastic units are usually not as rugged as comparable metal devices. All of the low power units just discussed have parallel leads which are closely spaced. These components are designed so that their leads can be inserted through the eyelets in printed circuit boards and then soldered. However, the leads can also be shortened so that these devices can be plugged into matching sockets or when necessary the leads can simply be spread apart and soldered to adjacent components.

Packages E and F are designed for high power applications. These packages are made of metal and are larger than the low power units previously described. The transistors are usually connected within these packages so that their collectors are securely attached to the metal cases. The heavy cases are then able to extract or drain away the heat that is generated within the transistor. These packages are also designed so that they can be securely fastened to a metal chassis or frame which can in turn extract the heat which accumulates within them. The external chassis or frame therefore, serves as a *heat sink* and effectively increases the power handling ability of the device. The two pins protruding from the bottom of package E serve as the emitter and base connections while the package case serves as the collector connection. The case of package F serves as the collector connection, however an additional collector lead (the hooked one) is also provided at the top of the package along with the emitter and base leads.

Transistor packages are usually referred to by a reference designation that indicates its size and configuration. The letters TO (meaning transistor outline) followed by a number is the commonly used package identifier. The common packages are TO-5 (Figure 5-5A), TO-18 (Figure 5-5B), TO-39 (Figure 5-5C), TO-92 (Figure 5-5D), TO-3 or TO-66 (Figure 5-5E) and TO-59 or TO-63 (Figure 5-5F).

Although all of the packages shown in Figure 5-5 are used to support, protect, and improve the operation of transistors, these same packages may also be used with other types of solid-state components. In fact many of the solid-state components that you will examine in the following units are mounted in packages which are either identical or similar to the ones shown here. In addition to these packages, there are many other types that are commonly used.

Due to the tremendous assortment of packages available, it is difficult to form absolute rules for identifying the various leads or terminals on each device. Each type of package has its particular lead arrangement and it is usually best to refer to the manufacturer's applicable specifications to identify the leads or terminals on each device.

## Self Test Review

1. A bipolar transistor which utilizes an N-type region which is sandwiched between two P-type regions is referred to as a __________ transistor.
2. A bipolar transistor which utilizes a P-type region which is sandwiched between two N-type regions is referred to as an __________ transistor.
3. Bipolar transistors have only one PN junction.
   a. True
   b. False
4. Bipolar transistors are formed by simply pressing together three alternately doped layers of germanium or silicon.
   a. True
   b. False
5. Because its diffused regions lie in a single plane, the transistor shown in Figure 5-3B is referred to as a __________ transistor.
6. Since the emitter and base regions of the transistor shown in Figure 5-4 rest on a plateau, the device utilizes a __________ type of construction.
7. Packages A, B, C, and D in Figure 5-5 are designed for low power applications.
   a. True
   b. False
8. The collectors of the transistors mounted in packages E and F of Figure 5-5 are connected to the package cases so that the __________ generated within the device may be quickly removed.

### Answers

1. PNP
2. NPN
3. b. False — Bipolar transistors have two PN junctions.
4. b. False — Bipolar transistors are formed by using the alloyed method, the diffusion method, the epitaxial growth method or by using other related and more refined methods of construction.
5. planar
6. mesa
7. a. True
8. heat

# BASIC TRANSISTOR ACTION

Although bipolar transistors may be used in various ways, their basic and most important function is to provide amplification. In other words bipolar transistors are used primarily to boost the strength or amplitude of electronic signals. The signals applied to a transistor and the amplified signals obtained from the transistor can be expressed in terms of voltage, current, or power. However, you will soon discover that the bipolar transistor is actually controlled by the current flowing through its terminals and this current is, in turn, controlled by varying the input voltage. Furthermore, the transistor must be properly biased by external voltages so that its emitter, base, and collector regions interact in the desired manner.

## Biasing NPN Transistors

Like the PN junction diode that you examined in a previous unit, the bipolar transistor must be properly biased in order to perform a useful function. However, the bipolar transistor has two PN junctions instead of just one like the diode and both of these junctions must be properly biased if the device is to function properly. For example, consider the NPN transistor shown in Figure 5-6. Notice than an external voltage has been applied to the base and collector regions. The junction formed between these two regions is commonly referred to as the *collector-base junction* or simply the *collector junction.* The collector and base regions effectively form a PN junction diode even though the base is very thin and lightly doped. The external voltage is being used to reverse bias the collection junction as shown. The reverse-biased collector junction will function in a manner similar to a reverse-biased PN junction diode and conduct only a very small leakage current. This extremely small current is supported by minority carriers in the N-type collector and P-type base regions. As explained in a previous unti, the minority carriers are the holes in the N-type material and the electrons in the P-type material. In this case the number of minority carriers in both regions (the base in particular) is very small and the resulting leakage current is therefore quite small. Under normal operating conditions , the collector junction of an NPN transistor is reverse-biased in this manner.

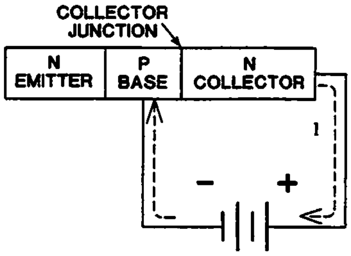


Figure 5-6
Reverse-biased collector junction

Now we will remove the external voltage shown in Figure 5-6 and apply it to the emitter and base regions of the NPN transistor as shown in Figure 5-7. Although the base region is extremely thin and lightly doped with respect to the emitter, these two regions still function exactly like a PN junction diode. The junction between these two regions is often referred to as the *emitter-base junction* or simply the *emitter junction*. The external voltage is being used to forward bias the

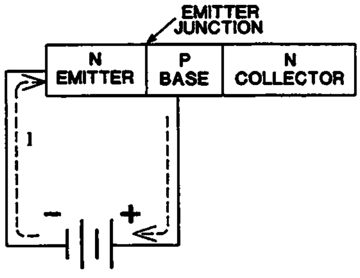


Figure 5-7
Forward-biased emitter junction

emitter junction because the N-type emitter and P-type base regions are respectively connected to the negative and positive terminals of the voltage source. The majority carriers within the two regions are now forced to combine at the emitter junction; however, the majority carriers (electrons) within the emitter greatly exceed the majority carriers (holes) in the base. This means that the base cannot supply enough holes to combine with all of the electrons that cross the junction. However, a certain amount of forward current will still flow through the emitter junction and the external voltage source. Under normal operating conditions the emitter junction must be forward-biased in this manner.

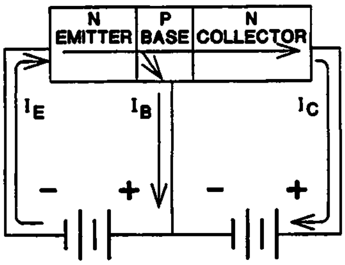


Figure 5-8
A properly biased NPN transistor showing internal and external current flow.

We have now seen the action which takes place in an NPN transistor when the collector junction and the emitter junction of the device are independently subjected to the proper bias voltages. However, our analysis of transistor action is not yet complete. This is because both junctions of the transistor must be simultaneously biased if the device is to operate properly. In other words, the emitter junction must be forward-biased and the collector junction must be reversed-biased as shown in Figure 5-8. When both junctions are biased in this manner, the action that takes place in the transistor is quite different.

We will now consider the action that takes place in the properly biased NPN transistor shown in Figure 5-8. Since the emitter junction is forward biased, electrons in the N-type emitter and the holes in the P-type base (the majority carriers) are forced to move toward the emitter junction and combine. However this portion of the transistor can no longer function like an independent diode and conduct a forward current which flows only through the emitter junction and its external bias voltage source. In fact, most of the current that flows into the emitter of the transistor will now flow completely through the transistor's base and collector regions. To understand how this action occurs we must consider both the physical and electrical characteristics of the transistor's emitter, base and collector regions.

The transistor has a very thin base region which is lightly doped with respect to the emitter. This means that the majority carriers (electrons in an NPN transistor) in the emitter will greatly exceed the majority carriers (holes) in the base. Therefore, most of the electrons that cross the emitter junction do not combine with holes and tend to accumulate in the base region. However, the electrons that are inserted or injected into the narrow base region by the emitter are now influenced by the positive potential applied to the collector region. In fact most of these

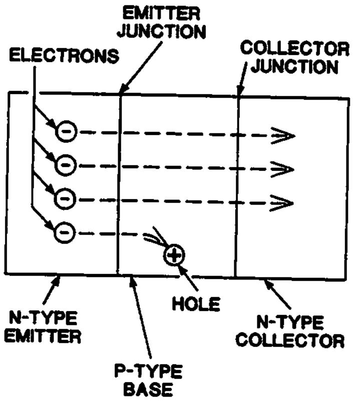


Figure 5-9
NPN transistor action

electrons are swept across the collector junction and into the collector region and from there into the positive side of the external voltage source which is used to reverse bias the collector junction. This action is shown in Figure 5-9. Typically, 95 to 99 percent of the electrons supplied by the emitter flow through the collector region and into the external voltage source. This current is referred to as *collector current* and is usually designated as $I_C$ as shown in Figure 5-8. The remaining emitter injected electrons (1 to 5 percent) combine with holes in the base region and therefore support a small current which flows out of the base region. This relatively small current is generally referred to as *base current* and is designated as $I_B$. The current that flows into the emitter region is generally referred to as *emitter current* and is designated as $I_E$. The relationship between $I_E$, $I_B$, and $I_C$ can be stated mathematically with the following equation:

$$I_E = I_B + I_C$$

The equation simply states that the emitter current ($I_E$) is equal to the sum of the base current ($I_B$) and the collector current ($I_C$). Another way of looking at this is that the collector current is equal to the emitter current less the current lost to the base.

$$I_C = I_E - I_B$$

If the base region of the transistor was not extremely thin, the action just described could not occur. The thin base region makes it possible for the emitter injected electrons to move quickly into the collector region. A large base region would minimize the interaction between the emitter and collector regions and the transistor would act more like two seperately biased diodes as shown in Figures 5-6 and 5-7. Furthermore, the emitter is doped much more heavily than the base region so that a large number of electrons will be supplied to the base and be subsequently pulled into the collector region. The collector region is lightly doped although not as lightly doped as the base region and the collector is also considerably larger than either the base or emitter regions. This relationship allows the collector to produce a sufficient number of minority carriers while the number of majority carriers are reduced to a low level. A large number of majority carriers could actually interfere with collector operation because they tend to inhibit the production of minority carriers.

In our analysis of transistor action we must also consider two additional points which are very important. First, it is important to realize that the emitter-base junction of the transistor has characteristics similar to those of a PN junction diode. A barrier voltage is therefore produced across the emitter junction of the device and this voltage must be exceeded before forward current can flow through the junction. Throughout our previous discussion we have assumed that the forward bias voltage exceeds this internal barrier voltage and we have also assumed that the emitter current is held to a safe value. The internal barrier voltage within a particular transistor is determined by the type of semiconductor material used to construct the device. The internal barrier voltage is typically 0.3 volts for germanium transistors and 0.7 volts for silicon transistors. This means that transistors, like PN junction diodes, will exhibit a relatively low voltage drop (0.3 or 0.7 volts) across their emitter junctions under normal operating conditions. A second important point to consider is that the collector-base junction of the transistor must be subjected to a positive potential that is high enough to attract most of the electrons supplied by the emitter. Therefore, the reverse bias voltage applied to the collector-base junction is usually much higher than the forward bias voltage across the emitter-base junction.

## Biasing PNP Transistors

Up to this point we have considered only the action which takes place in a properly biased NPN transistor. The only difference between the NPN and PNP transistors is the type of charge carriers involved and the polarity of the external bias voltages. The NPN transistor utilizes electrons as charge carriers but the PNP transistor utilizes holes. In other words, NPN transistors depend on the movement of electrons between their emitter, base, and collector regions while the operation of PNP devices is explained by the movement of holes. Also, since the structures of PNP and NPN transistors are opposite, these devices require external bias voltages which are also opposite. This means that the external bias voltages shown in Figure 5-8 must be reversed so that the PNP transistor is biased as shown in Figure 5-10. If you examine this figure closely you will see that the emitter junction is forward-biased and the collector junction is reverse-biased. Figure 5-10 also shows that the emitter, base, and collector currents flowing through the PNP device are opposite to those flowing through the NPN device.

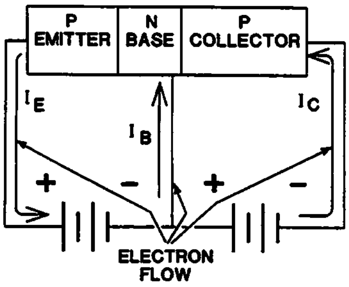


**Figure 5-10**
A properly biased PNP transistor showing external current flow.

## Self Test Review

9. A bipolar transistor has ________ PN junctions.
10. The junction formed between the emitter and base regions is called the ____________ junction.
11. The ____________ junction in a bipolar transistor is usually reverse-biased.
12. The ____________ junction in a bipolar transistor is usually forward-biased.
13. The collector current flowing through the NPN transistor shown in Figure 5-8 is exactly equal to the emitter current.

    a. True

    b. False
14. The base current in a bipolar transistor will usually be equal to only 1 to 5 percent of the emitter current.

    a. True

    b. False

### Answers

9. two
10. emitter-base or emitter
11. collector
12. emitter
13. b. False — the collector current is always slightly less than the emitter current.
14. a. True

# TRANSISTOR AMPLIFICATION

Now that we have seen the basic action that takes place in a properly biased transistor we will go one step further and see how a transistor can be used to amplify electronic signals. We will continue our analysis of the NPN transistor shown in Figure 5-8; however, we will replace the transistor with its electronic symbol and therefore, represent the circuit as shown in Figure 5-11. Notice that the external voltage sources that are used to forward bias the emitter junction and reverse bias the collector junction have also been labeled $V_{EE}$ and $V_{CC}$ respectively. These two symbols are widely accepted as standard designations for the operating voltages used in this type of circuit. Furthermore, we have made $V_{EE}$ variable so that we may vary the forward bias voltage applied to the emitter junction. The transistor has also been designated as $Q_1$ since the Q designation is widely used to represent transistors in electronic circuits.

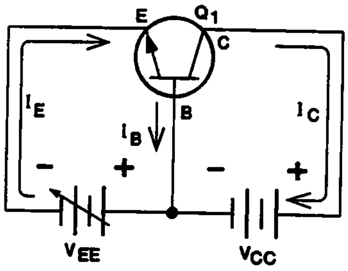


Figure 5-11
Schematic representation
of a properly biased NPN transistor

In the previous discussion on transistor action we saw that the emitter current ($I_E$), base current ($I_B$) and collector current ($I_C$) could be related mathematically by the expression $I_E = I_B + I_C$ and we assumed that all three currents were constant. However, all three of these currents vary in a proportional manner as indicated by the mathematical expression previously stated. For example, if $I_E$ doubled in value, then $I_B$ and $I_C$ would likewise double in value. The reason for this is quite simple and can be easily explained. The number of electrons that the emitter supplies to the base region is determined by the amount of forward bias voltage across the emitter junction. When the forward bias voltage increases ($V_{EE}$ increases) more electrons are injected into the base region ($I_E$ increases), therefore more electrons are swept into the collector region to become collector current ($I_C$ increases). The additional electrons that do not move into the collector region combine with holes in the base region and become base current ($I_B$ increases). When the forward bias voltage decreases, the opposite effect is produced and all three currents decrease. The three currents tend to be directly proportional or in other words they all vary by the same percentage.

In order for the transistor shown in Figure 5-11 to provide amplification, the device must be capable of accepting an input signal (current or voltage) and providing an output signal that is greater in strength or amplitude. The transistor cannot perform this function if it remains connected as shown in Figure 5-11. Instead we must make several changes to the circuit so that it appears as shown in Figure 5-12. Notice that we have added a resistor between the collector of transistor $Q_1$ and the positive terminal of $V_{CC}$. This component is commonly referred to as a load resistor and is labeled $R_L$. This resistor is used to develop an output voltage with the polarity indicated. In other words the collector current must flow through this resistor to produce a specific voltage drop. We have also separated the connection between the emitter of transistor $Q_1$ and the negative terminal of $V_{EE}$. These open connections serve as the input of the amplifier circuit and allow an input signal to be inserted between the emitter of $Q_1$ and $V_{EE}$. Therefore the input is not an open circuit, but is completed when an external voltage source is connected between the emitter and $V_{EE}$. Furthermore, the values of $V_{EE}$ and $V_{CC}$ have been selected to bias transistor $Q_1$ so that the values of $I_E$, $I_B$ and $I_C$ are high enough to permit proper circuit operation.

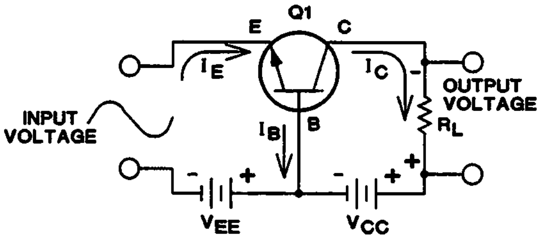


**Figure 5-12**
A basic NPN transistor amplifier circuit.

The emitter junction of transistor $Q_1$ is forward-biased and like a PN junction diode, has a relatively low resistance. However, the collector junction is reverse-biased and therefore has a relatively high resistance. In spite of the tremendous difference in emitter junction resistance and collector junction resistance the emitter current ($I_E$) is almost equal to the collector current ($I_C$). The value of $I_C$ is only slightly less than $I_E$ because a small portion of $I_E$ flows out of the base region to become base current ($I_B$). The load resistor can therefore have a high value without greatly restricting the value of $I_C$.

If a small voltage is applied to the input terminals which aids the forward bias voltage ($V_{EE}$), the value of $V_{EE}$ is effectively increased. This will cause $I_E$ to increase and $I_C$ will also increase. This will increase the voltage drop across the load resistor ($R_L$). However, when the input voltage is reversed so that it opposes $V_{EE}$, the value of $V_{EE}$ is effectively reduced and both $I_E$ and $I_C$ will decrease. This will cause the voltage drop across $R_L$ to decrease. A change in input voltage can therefore produce a corresponding change in output voltage, however the output voltage change will be much greater than the input change. This is because the output voltage is developed across a high output load resistance ($R_L$) and the input voltage is applied to the low resistance offered by the emitter junction. Therefore, a very low input voltage can control the value of $I_E$ which in turn determines the value of $I_C$. Even though $I_C$ remains slightly less than $I_E$, it is forced to flow through a higher resistance and therefore produce a higher output voltage.

The action just described will occur if the input signal is either a dc or an ac voltage. The important point to remember is that any change in input voltage is greatly amplified by the circuit so that a much larger but proportional change in output voltage is obtained. It is also important to realize that the input signal is not simply increased in size, but is used to control the conduction of transistor $Q_1$ and this transistor in turn controls the current through the load resistor ($R_L$). The transistor is therefore made to take energy from the external power source $V_{CC}$ and apply it to the load resistor in the form of an output voltage whose value is controlled by a small input voltage.

The transistor shown in Figure 5-12 is therefore being used to convert a relatively low voltage to a much higher voltage. Any circuit which performs this basic function is commonly referred to as a *voltage amplifier*. Such circuits are widely used in electronic equipment since it is often necessary to raise both dc and ac voltages to higher levels so that they may be effectively used. Later in this unit you will also discover that the transistor can be connected in a different manner which will allow it to amplify currents as well as voltages. Any circuit which is used for the sole purpose of converting a low current to a higher current is commonly referred to as a *current amplifier*.

Although an NPN transistor was used in the amplifier circuit shown in Figure 5-12, a PNP transistor will perform the same basic function. However; in the PNP circuit the bias voltages ($V_{EE}$ and $V_{CC}$) must be reversed and the respective currents ($I_E$, $I_B$, and $I_C$) will flow in the opposite directions.

# Self Test Review

15. The emitter, base, and collector currents in a bipolar transistor tend to vary by the same percentage.
    a. True
    b. False
16. PNP transistor bias polarities are applied just like an NPN transistor.
    a. True
    b. False
17. The transistor shown in Figure 5-12 is being used as a __________ amplifier.
18. If an input voltage is applied to the circuit shown in Figure 5-12 which aids the forward bias voltage ($V_{EE}$), the voltage across $R_L$ will __________.
19. Only NPN transistors may be used in the amplifier circuit shown in Figure 5-12, regardless of which direction $V_{EE}$ and $V_{CC}$ are connected.
    a. True
    b. False

## Answers

15. a. True
16. b. False — The PNP transistor requires bias voltages that have polarities opposite to those used with the NPN transistor.
17. voltage
18. increase
19. b. False — Both NPN and PNP transistors may be used in the circuit, assuming that the correct polarity voltages are used.

# TRANSISTOR CIRCUIT ARRANGEMENTS

As explained previously, the bipolar transistor is used primarily as an amplifying device. However, there is more than one way of using a transistor to provide amplification and each method offers certain advantages or benefits but also has certain disadvantages or limitations. Basically, the bipolar transistor may be utilized in three different circuit arrangements to perform its important amplifying function. In each arrangement one of the transistor's three leads is used as a common reference point and the other two leads serve as input and output connections. The three circuit arrangements are referred to as the **common-base circuit**, the **common-emitter circuit**, and the **common-collector circuit**, and each arrangement may be constructed by using either NPN or PNP transistors. In many cases the common leads are connected to circuit or chassis ground and the circuit arrangements are then often referred to as grounded-bases, grounded-emitter, and grounded-collector circuits. Furthermore, in each circuit arrangement the transistor's emitter junction is always forward-biased while the collector junction is reverse-biased. We will now briefly examine each of these basic circuit arrangements so that the flexibility of the bipolar transistor can be made apparent. In the next unit of instruction these circuits will again be described but in much greater detail.

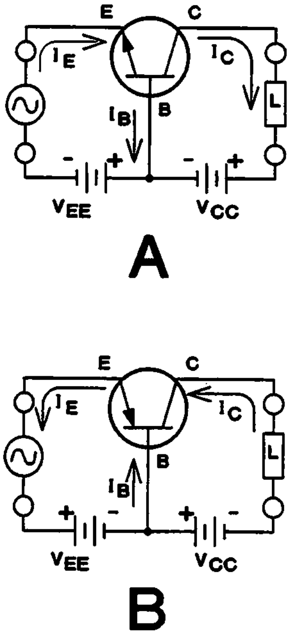


Figure 5-13
Common-base circuits

## Common Base Circuits

In the common-base circuit, the transistor's base region is used as a common reference point and the emitter and collector regions serve as the input and output connections. This basic circuit arrangement is shown in Figure 5-13. Figure 5-13A shows how an NPN transistor is connected in the common-base configuration while Figure 5-13B shows how the same arrangement is formed with a PNP transistor. Notice that both circuits are arranged in the same basic manner, but the polarity of the bias voltages ($V_{EE}$ and $V_{CC}$) are opposite so that the NPN and PNP transistors are properly biased. This common-base arrangement was also described in the previous sections of this unit although the previous circuits were not identified as common-base circuits. This arrangement was used in the previous discussions because it serves as an excellent model for explaining basic transistor operation.

In the common-base circuit the input signal is applied between the transistor's emitter and base and the output signal appears between the transistor's collector and base. The base is therefore common to both the input and output. The input signal applied to the circuit is represented by the ac generator symbol. The output voltage is usually developed across a resistive component but in some cases components which have inductive as well as resistive properties (coils, relays, or motors) may be used. In general, the component that is connected to the output of the circuit is referred to as a *load*. In Figure 5-13 the load is identified by the box which contains the letter "L".

The common-base circuit is useful because (as explained earlier) it provides voltage amplification. However, this circuit has other basic characteristics which should be considered. Since the emitter junction of the transistor is forward-biased the input signal sees a very low emitter-to-base resistance or in other words the transistor has a low input resistance. However, the collector junction of the transistor is reverse-biased and therefore offers a relatively high resistance. The resistance across the output terminals (collector-to-base) of the transistor is therefore much higher than the input resistance. The input resistance of a typical low power transistor that is connected in the common-base configuration could be as low as 30 or 40 ohms; however, the output resistance of the transistor could be as high as 1 megohm.

The common-base circuit (as explained previously) provides slightly less output current than input current because $I_C$ is slightly lower than $I_E$. In other words there is a slight current loss between the input and output terminals of the circuit. However, the circuit is still useful as a voltage amplifier as explained earlier. Since the input and output currents are almost the same and because the output voltage developed across the load can be much higher than the input signal voltage, the output power produced by the circuit is much higher than the input power applied to the circuit. The common-base circuit therefore provides power amplification as well as voltage amplification.

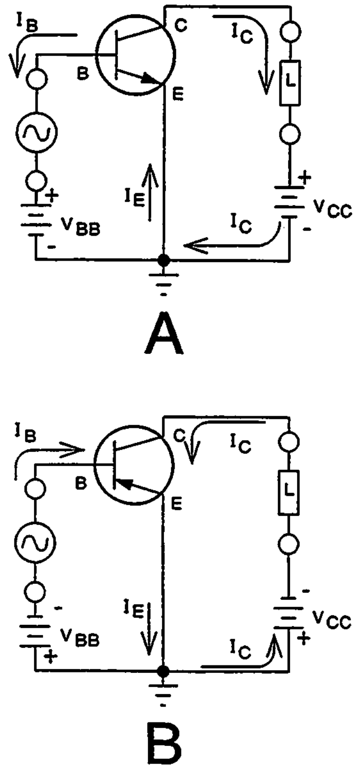


Figure 5-14
Common-emitter circuits

## Common Emitter Circuits

Bipolar transistors (NPN and PNP) may also be connected as shown in Figure 5-14. Notice that the input signal is applied between the base and emitter of each transistor while the output signal is developed between the collector and emitter of each transistor. Since the emitter of each transistor is common to the input and output of each circuit, the arrangements are referred to as common-emitter circuits. Notice that the emitter junction of each transistor is still forward-biased while the collector junction of each device is reverse-biased. Also, the NPN circuit shown in Figure 5-14A and the PNP circuit shown in Figure 5-14B require bias voltages that are opposite in polarity. Notice that the forward bias voltage (designated as $V_{BB}$) is now applied to the base of the transistor instead of the emitter. In other words the forward bias voltage ($V_{BB}$) is applied so that it controls the base current ($I_B$) instead of the emitter current ($I_E$). Only the base current flows through bias voltage

$V_{BB}$ and the ac generator (signal source). The reverse bias voltage ($V_{CC}$) is applied through the output load to the collector and emitter regions of the transistor and only the collector current ($I_C$) flows through the load and $V_{CC}$.

Although it may not be apparent, $V_{CC}$ does actually reverse bias the collector-base junction of the transistor. Figure 5-15 shows how this reverse bias is obtained in an NPN common-emitter circuit. If $V_{CC}$ is connected across the emitter and collector regions of the transistor as shown, the emitter-base junction becomes forward-biased and therefore has a low resistance. However, $V_{CC}$ causes the N-type collector to be positive with respect to the P-type base and the collector-base junction is therefore reverse-biased. The forward-biased emitter-base junction effectively allows the negative terminal of $V_{CC}$ to be applied to the P-type base region. $V_{CC}$ is therefore effectively applied to the collector and base regions as shown. If $V_{CC}$ was reversed, the collector junction would become forward-biased and the emitter junction would become reverse-biased; however, this condition would be undesirable. The same basic action also occurs in a PNP transistor even though the PNP device requires bias voltage polarities which are exactly opposite to those used with an NPN device.

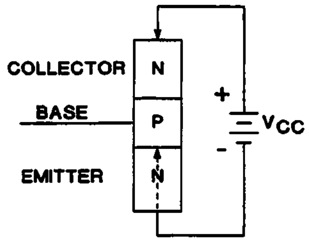


Figure 5-15
The reverse-biased collector junction in an NPN common-emitter circuit.

The three currents ($I_E$, $I_B$, and $I_C$) in the common-emitter circuit still have the same relationship as before and still vary in a proportional manner. In other words $I_E$ is still equal to $I_B + I_C$ even though $I_B$ has now become the input current and $I_C$ has become the output current. Any change in the input current ($I_B$) will therefore result in a proportional change in $I_E$ and $I_C$. The difference between this circuit and the common-base circuit previously described is therefore simply one of reference. We have now connected the transistor so that its very low base current is being used to control its much larger collector current.

When an input signal voltage is applied to the common-emitter circuit, it either aids or opposes $V_{BB}$ and therefore causes $I_B$ to either increase or decrease. The much higher emitter and collector currents (which are almost equal) are forced to increase and decrease by the same percentage. Since a very low $I_B$ is used to control a relatively high $I_C$, the transistor is now being used to provide current amplification as well as voltage amplification. This of course means that the power supplied to the load is much greater than the power applied to the input of the circuit. In fact the common-emitter arrangement provides the highest power output for a given power input when compared with the other two circuit arrangements. The common-emitter circuit is also the only circuit that provides current, voltage, and power amplification.

A transistor that is used in a common-emitter arrangement will also have a low input resistance but this resistance will not be as low as the input resistance of a common-base arrangement. This is because the input base current for the common-emitter circuit is much lower than the input emitter current of the common-base circuit. Also, the output resistance of a transistor connected in a common-emitter arrangement will be somewhat lower than when it is connected in a common-base circuit. A typical low power transistor might have an input resistance of 1000 or 2000 ohms and an output resistance of 50,000 or 60,000 ohms.

## Common Collector Circuits

The NPN and PNP transistors shown in Figure 5-16 are connected so that the collector of each transistor is used as a common reference point. The base and the emitter regions serve as the input and output connections. These circuits are known as common-collector circuits.

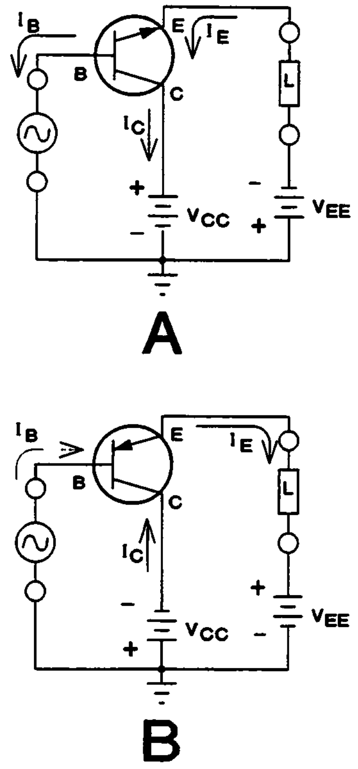


Figure 5-16
Common-collector circuits

In the common-collector circuit the input signal voltage is applied between the base and collector regions of the transistor. The input either aids or opposes the transistor's forward bias and therefore causes $I_B$ to vary accordingly. This in turn causes $I_E$ and $I_C$ to vary by the same percentage. An output voltage is developed across the load which is connected between the emitter and collector regions of the transistor. The emitter-current ($I_E$) flowing through the load is much greater than the base current ($I_B$) and the circuit therefore provides an increase in current between its input and output terminals. However, the voltage developed across the load will always be slightly lower than the voltage applied to the circuit. The slightly lower voltage will appear at the emitter of the transistor because the device tends to maintain a relatively constant voltage drop across its emitter-base junction. As explained earlier, this forward voltage drop may be equal to approximately 0.3 volts if the transistor is made of germanium or 0.7 volts if the transistor is made of silicon. The output voltage appearing at the emitter of the transistor therefore tends to track or follow the input voltage applied to the transistor's base. For this reason the common-collector circuit is often called an *emitter follower*.

The common-collector circuit therefore functions as a current amplifier but does not produce an increase in voltage. However, the increase in output current (even though output voltage is slightly less than input voltage) results in a moderate increase in power. Also, the input resistance of any transistor connected in a common-collector arrangement is extremely high. This is because the input resistance is that resistance which appears across the reverse-biased collector-base junction. This input resistance can be as high as several hundred thousand ohms in a typical low power transistor. However, the output resistance appearing between the transistor's emitter and collector regions will be much lower (often as low as several hundred ohms) because of the relatively high emitter current ($I_E$) that flows through the output lead.

The common-collector circuit is not useful as an amplifying device. Instead, this circuit is widely used in applications where its high input resistance and low output resistance can perform a useful function. The circuit is often used to couple high impedance sources to low impedance loads and therefore perform the same basic function as an impedance matching transformer. The current does provide a substantial amount of power gain.

The bias voltages ($V_{BB}$, $V_{EE}$, and $V_{CC}$) used in the three circuit arrangements just described may appear to restrict or interfere with the flow of signal current through these circuits; however, this does not occur. The bias voltages are effectively shorted as far as the input and output signals are concerned, although in some applications large capacitors are placed across the voltage sources to insure that they offer minimum impedance to signal currents. For example, a large electrolytic capacitor could be placed across $V_{CC}$ in Figure 5-16 to insure that the collector of the transistor is at ground potential as far as ac signals are concerned.

## Self Test Review

20. A transistor that is connected in a common-base arrangement provides a substantial current gain.

    a. True

    b. False

21. Regardless of the type of circuit arrangement used, the transistor's emitter junction must always be __________-biased.

22. The component connected to the output of a common-base, common-emitter, or common-collector circuit is commonly referred to as a __________.

23. The common-base circuit provides voltage and power amplification.

    a. True

    b. False

24. The transistor used in a common-base circuit will have a low input __________.

25. The common-emitter circuit provides an increase in current, voltage, and power.

    a. True

    b. False

26. In a common-emitter circuit $I_B$ serves as the __________ current.

27. The common-__________ circuit arrangement provides the highest power amplification.

28. In the common-collector circuit, the input signal is applied between the __________ and __________ regions of the transistor.

29. The common-collector circuit does not provide voltage amplification.

    a. True

    b. False

## Answers

20. b. False — The output collector current is slightly less than the input emitter current in a common-base circuit.
21. forward
22. load
23. a. True
24. resistance
25. a. True
26. input
27. emitter
28. base and collector
29. a. True

# TESTING BIPOLAR TRANSISTORS

Although bipolar transistors are solid-state devices which are capable of operating for extremely long periods of time without failure, these devices do occasionally become defective. The failure of a particular transistor can be caused by excessively high temperatures, current, voltages or by simply subjecting the component to extreme mechanical stress. As a result of these electrical or mechanical abuses, the transistor may short or open internally or in some cases the characteristics of the device may be altered.

## Test Procedures

Test equipment is readily available for testing transistors either in or out of an electronic circuit. These instruments will indicate when a transistor is open or shorted, how well it amplifies and whether or not the device is passing an excessive amount of undesirable leakage current. A simple transistor tester is shown in Figure 5-17. However, the most common troubles (opens and shorts) can be easily detected with an ohmmeter.

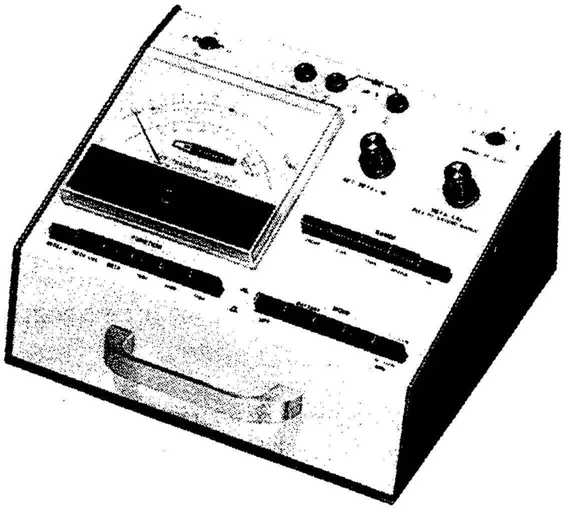


**Figure 5-17**
A simple bipolar transistor and junction
diode tester that checks for shorts, opens, gain and leakage.

The ohmmeter may be used to determine if a short or open exists between the transistor's emitter and base, base and collector, or emitter and collector. When checking a transistor with an ohmmeter, you are simply looking for unusually low (near zero) or unusually high (nearly infinite) resistances. Since a bipolar transistor has two PN junctions, it may be compared with two diodes that are connected back-to-back. Each diode (PN junction) will exhibit a low resistance when forward-biased and a high resistance when reverse-biased. The ohmmeter battery is the source of the forward and reverse bias voltage. In order to test a transistor we must therefore check the forward and reverse resistance of each junction.

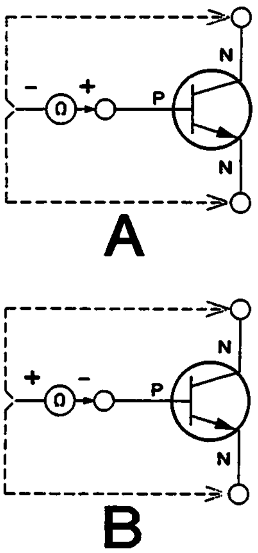


Figure 5-18
Testing an NPN transistor

## Testing NPN Transistors

To check the forward resistance of each junction in an NPN transistor, the ohmmeter leads must be connected as shown in Figure 5-18A. The positive (normally red) lead is connected to the base while the negative (usually black) lead is connected first to the transistor's emitter and then to its collector. The ohmeter will send a forward current first through the emitter base junction, then through the base-collector junction and should indicate that both junctions have a relatively low resistance (typically several hundred ohms or less). Next, the ohmmeter leads are reversed and both junctions are again checked as shown in Figure 5-18B. This time both junctions are reverse-biased and the ohmmeter should indicate that both junctions have a relatively high resistance (several hundred thousand ohms or greater).

## Testing PNP Transistors

When checking a PNP transistor the same basic procedure is followed; however, the ohmmeter connections are exactly opposite to those shown in Figure 5-18. The proper ohmmeter connections for checking the forward resistance of the two junctions in a PNP device are shown in Figure 5-19A and the proper connections for checking reverse resistance is shown in Figure 5-19B.

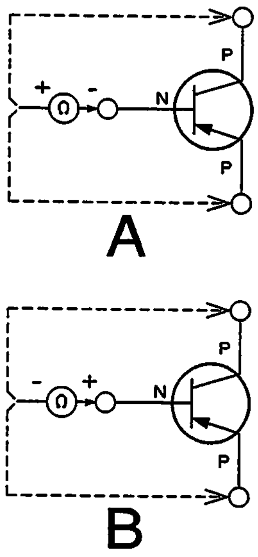


Figure 5-19
Testing a PNP transistor

Although the two test procedures described above are quite simple, several general precautions should be observed when the tests are performed. When checking the forward resistance of the junctions, you should use a low resistance range that will allow the meter to present a mid-range indication or one that is easy to read. When checking the reverse resistance, a higher range must be used to obtain a convenient reading. Furthermore, the specific resistance readings that you obtain have no real meaning. You are interested only in determining that each PN junction has a high ratio of forward and reverse resistance. However, through experience you will find that these forward and reverse resistances generally fall within certain ranges. Several hundred ohms of forward resistance and several hundred thousand ohms to several megohms of reverse resistance is typical for both low and medium power transistors, but high power transistors generally have forward and reverse resistances that are somewhat lower. Also, transistors that are made from silicon, usually have higher forward and reverse resistances than transistors that are made from germanium.

When checking low power transistors you should avoid using the lowest or highest resistance ranges on your ohmmeter if possible, because your ohmmeter will supply its maximum current or voltage to the transistor when these ranges are used. In most cases this will not damage a transistor but it is still a good practice.

When a junction has very low forward and reverse resistances (particularly when both resistances are equal) the junction is effectively shorted and the transistor is therefore defective. When a junction has an extremely high resistance in both the forward and reverse directions, it is effectively open and the transistor is again considered to be defective. Both the forward and reverse junction resistances must be measured before the true condition of the junction is revealed.

When you check transistors with an ohmmeter, you may notice that the forward and reverse resistances change when different resistance ranges are used. This occurs because the ohmmeter applies a different voltage and current to the transistor on each resistance range. In addition, resistance readings often vary considerably from one ohmmeter to another. One way of increasing the reliability of PN junction resistance measurements is to compare the resistance readings of the transistor under test with resistance readings on a known good transistor of the same type.

## Identifying A Transistor

Sometimes you may find it necessary to determine if a given transistor is an NPN or PNP device. This identification can also be performed with an ohmmeter. As in testing a transistor with an ohmmeter, you must know the polarity of the voltage at the ohmmeter leads. Normally, the internal ohmmeter battery will be connected so that a positive potential will appear at the red or plus lead of the ohmmeter and a negative potential appears at the black or minus lead. In some ohmmeters this is not true, however. Check your unit by referring to the ohmmeter circuit.

In order to forward bias a PN junction, you must apply a bias voltage to it so that the cathode (N) is negative and the anode (P) is positive. You must reverse the polarity to reverse bias the junction. With this information and the knowledge of your ohmmeter operation, you can identify a transistor as being PNP or NPN. The procedure below tells you how.

1. Identify the transistor leads. Locate the emitter and base connections.
2. Set your ohmmeter to the RX 10 or RX 100 range.
3. Connect the ohmmeter to the emitter and base leads of the transistor. It doesn't matter which polarity you use.
4. Note the resistance reading. If it is low, the junction is forward biased. If it is high the junction is reversed biased.
5. Note the polarity of the connection of the ohmmeter to the emitter lead.
6. Refer to the table below which will identify the transistor.

| Resistance Reading | –Ohmmeter lead to emitter | +Ohmmeter lead to emitter |
|---|---|---|
| high | PNP | NPN |
| low | NPN | PNP |

As an example of how to use the table, consider the case of a high resistance reading when the ohmmeter negative (–) lead is connected to the emitter. The high resistance indicates reverse bias on the emitter-base junction. Therefore the negative (–) lead must be on the P section of the junction (the emitter) and the + lead on an N section (the base). Therefore, we have a PNP transistor.

## Self Test Review

30. Bipolar transistors may be quickly tested for opens or shorts by using an ____________.
31. In testing a bipolar transistor you may observe that the reverse resistance of the emitter-base junction is low. Therefore, the junction is:
    a. open
    b. shorted
    c. good
    d. can't tell
32. In testing a bipolar transistor you observe that the reverse resistance of the base-collector junction is high. Therefore, the junction is:
    a. open
    b. shorted
    c. good
    d. cannot be determined

### Answers

30. ohmmeter
31. b. shorted
32. d. cannot be determined — In order to determine whether a PN junction is good or bad, you must measure both the forward and reverse resistances. A single measurement will not always suffice. In this case a reverse resistance measurement on the base-collector junction is high. The forward resistance should be low. But it too could be high if the junction is open.

# EXPERIMENT 5

## Testing Bipolar Transistors

*OBJECTIVE:* *To learn how to test and identify NPN and PNP bipolar transistors.*

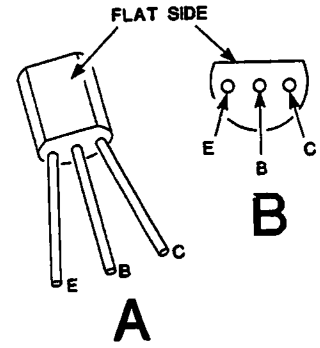


**Figure 5-20**
Locating the emitter, base, and collector leads NPN, 417-801.

### Introduction

In this experiment, you will use an ohmmeter to test an NPN and a PNP transistor. To perform this experiment you will need an ohmmeter that is capable of measuring resistances as low as several ohms and resistances as high as several megohms. A meter that has low resistance ranges of either R × 1 or R × 10 and high resistance ranges of either R × 100k or R × 1M is desirable.

### Material Required

Ohmmeter
1 — NPN silicon transistor MPSA20 (417-801)
1 — PNP transistor X29A829 (417-201)

NOTE: Many digital and electronic meters now provide a diode check mode, indicated by a diode symbol. In which, a constant current source instead of a constant voltage source is sometimes provided. A silicon diode will indicate between 0.5 and 0.7 $V_F$. A germanium diode should indicate between 0.2 and 0.4 $V_F$. In some cases, a diode is indicated, but the meter will still measure ohms. This meter position is designed to measure the forward and reverse resistances of a diode. If you are not sure how the function on your meter works, check the meters owner manual.

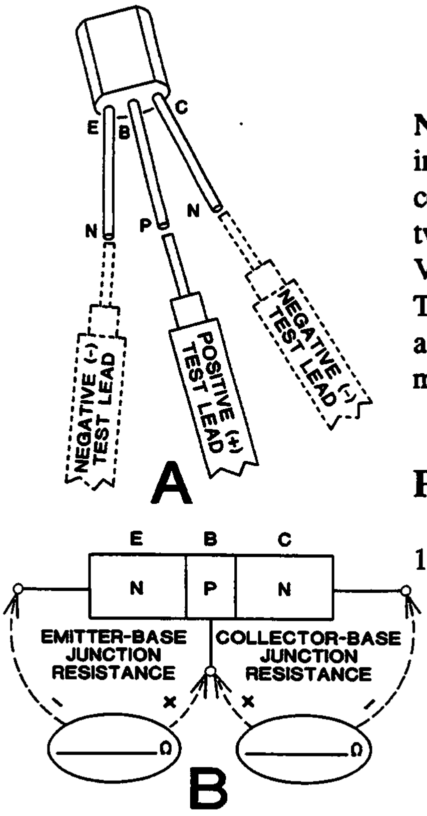


Figure 5-21
Checking the forward resistance of each junction in an NPN transistor.

### Procedure

1. Locate the NPN transistor and hold it (flat side up) as shown in Figure 5-20A. The transistor's emitter (E) lead will be to the left, the base (B) lead will be in the middle, and the collector (C) lead will be on the right side. Spread the leads apart, as shown, so that you will be able to connect your ohmmeter leads to them without shorting them together. Notice that an additional outline drawing of the transistor (as viewed from the bottom) is shown in Figure 5-20B. Manufacturers generally provide this type of outline drawing with the transistor's specifications to help in identifying the transistor's emitter, base, and collector leads.

2. Now you will use your ohmmeter to measure the forward resistance of the transistor's emitter-base and collector-base junctions. To do this you should use the ohmmmeter's R × 10 or R × 100 range. Then connect the ohmmeter's positive test lead (this lead is usually red) to the transistor's base lead as shown in Figure 5-21A. Then alternately connect the ohmmeter's negative test lead (usually black) to the transistor's emitter and collector leads. Note the forward resistance across each junction and record the two readings in the spaces provided in Figure 5-21B. Notice that the positive test lead must be connected to the P-type base while the negative test lead is connected to the N-type emitter or collector.

3. Now, you will use your ohmmeter to measure the reverse resistance of the transistor's emitter-base and collector-base junctions. Use the ohmmeter's R × 10k or R × 100k range, but this time, connect the ohmmeter's negative test lead to the base, and alternately connect the positive test lead to the emitter and collector, as shown in Figure 5-22A. Note the reverse resistance across each junction and record your readings in the space provided in Figure 5-22B.

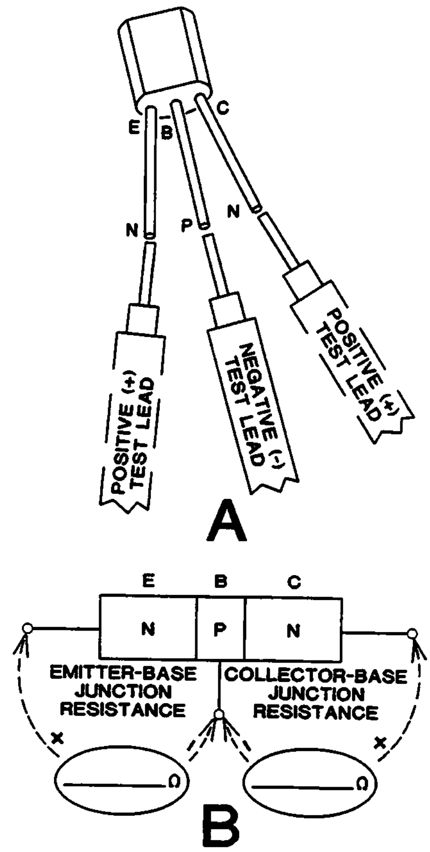


Figure 5-22
Checking the reverse resistance of each junction in an NPN transistor.

## Discussion

In this portion of the experiment you measured the forward and reverse resistance across the emitter-base and collector-base junctions of an NPN transistor. You should have discovered that the forward resistance of each junction was much lower than the reverse resistance, thus, indicating that each junction is functioning like a PN junction diode. You probably measured a forward resistance of several hundred ohms, or more, if you used the R × 100 range; or, your reading may have been much lower if you used the R × 10 range. However, your reverse resistance reading should have been extremely high (your meter probably indicated that the reverse resistance was infinite). The exact readings that you obtained are not as important as the difference between the two readings. One reading should be relatively low and the other very high if the transistor is functioning properly. Only when a particular junction has shorted or opened will it exhibit a very low or very high resistance in both directions. Since your NPN transistor is made from silicon, it will allow very little leakage current to flow through either junction that is reverse-biased. This is why your reverse resistance readings were so high.

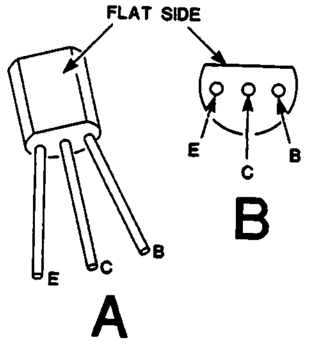


**Figure 5-23**
Locating the emitter, base and collector leads, PNP, 417-201.

## Procedure (continued)

4. Locate your PNP transistor and position the device as shown in Figure 5-23A. Notice the location of the emitter, base, and collector leads. These leads are also identified in the outline drawing shown in Figure 5-23B. Spread the leads apart as shown.

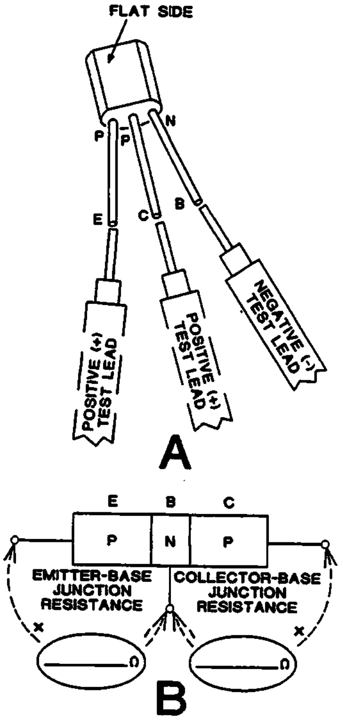


**Figure 5-24**
Checking the forward resistance of each junction in a PNP transistor.

5. Now you will use your ohmmeter to measure the forward resistance of the transistor's emitter-base and collector-base junctions. Use the ohmmeter's R × 10 or R × 100 range. Alternately connect the ohmmeter's test leads as shown in Figure 5-24A and record your forward resistance readings in the spaces provided in Figure 5-24B. Notice that the negative test lead must be connected to the N-type base while the positive test lead is connected to the P-type emitter or collector.

6. Now you will use your ohmmeter to measure the reverse resistance of the transistor's emitter-base and collector-base junctions. Use the ohmmeter's R × 10k or R × 100k range but this time connect the ohmmeter's test leads as shown in Figure 5-25A. Record your reverse resistance readings in the spaces provided in Figure 5-25B.

## Discussion

In this portion of the experiment you measured the forward and reverse resistances across the emitter-base and collector-base junctions of a PNP transistor. Your measurements should have been similar to the NPN transistor. If a germanium transistor had been tested, the reading would have been much lower. Germanium transistors always have lower junction resistances then similar types made from silicon.

## Procedure (continued)

7. Select either of the transistors that you used in the previous steps and without looking to see if it is the NPN or PNP type place the transistor as shown in Figure 5-26, on your work surface. Assume that the leads are numbered 1, 2, and 3 counting from left to right.

8. With your ohmmeter set to the R × 10k scale, connect the meter as shown in Figure 5-26. Record your measurements in the spaces provided in Figure 5-26. If your readings are extremely high, list them in the table as high. If the readings are mid scale or below, you may list them as either low, or the actual reading. Measure the 6 combinations shown in the figure and record the readings for each measurement.

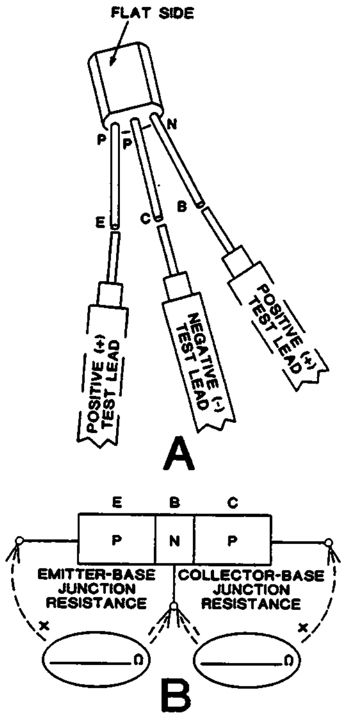
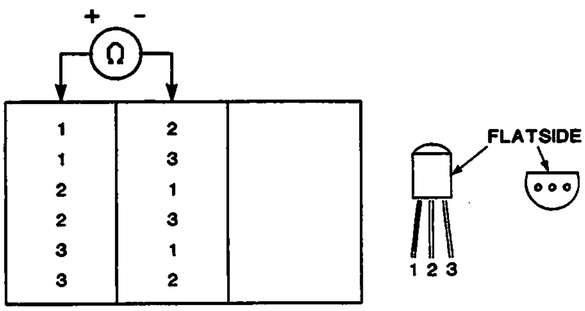


**Figure 5-25**
Checking the reverse resistance of each junction in a PNP transistor.

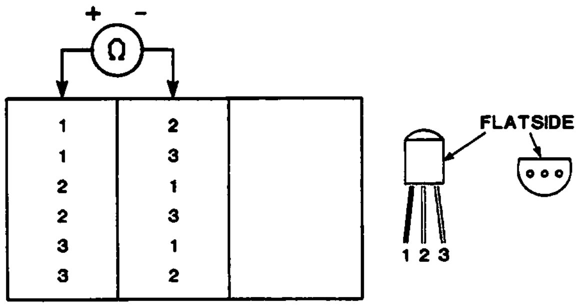


| | | |
|---|---|---|
| 1 | 2 | |
| 1 | 3 | |
| 2 | 1 | |
| 2 | 3 | |
| 3 | 1 | |
| 3 | 2 | |

**Figure 5-26**

Chart for recording data from step 8.

9. Now use the other transistor and repeat steps 7 and 8. Record your readings in the spaces provided in Figure 5-27.



| | | |
|---|---|---|
| 1 | 2 | |
| 1 | 3 | |
| 2 | 1 | |
| 2 | 3 | |
| 3 | 1 | |
| 3 | 2 | |

**Figure 5-27**

Chart for recording data from step 9.

## Discussion

In step 7, you selected an unknown transistor and without knowing which lead was the base or emitter or collector, you assigned the leads the numbers 1 through 3. Next, you measured the 6 possible ways to configure your ohmmeter and recorded those measurements in the spaces provided in Figure 5-26 as either a high or a low reading, or the actual value.

Examine the recorded data, you should see that there are two low readings and four high readings. The two low readings will both include one of the lead's numbers. This is the base lead of the transistor. Further examination of the data will reveal that the base lead had the same polarity in both cases. If the polarity applied to the base to obtain the low readings is negative, then the transistor's base is negative and it is a PNP transistor. If the two low readings were obtained using the meter's positive lead, then the base is positive and it is on NPN transistor.

Examine the data in Figure 5-27 and verify the information in the paragraph above. Now look at the transistors and determine if the 417-201 was verified as a PNP. The other transistor, 417-801, should have tested out to be an NPN. Remember, a forward biased junction has a lower resistance than a reverse biased junction. Also remember that the base of a transistor is always forward biased for conduction.

An ohmmeter can be used to test a transistor for shorted (the most common problem) or open junctions, and to establish if an unknown transistor is a PNP or NPN. It is also possible to determine the transistor's base lead.

# EXPERIMENT 6

## Transistor Characteristics

*OBJECTIVE:* *To examine the ability of the bipolar transistor as a switch.*

### Introduction

In this experiment, you will construct a simple NPN transistor circuit. In this case, a common emitter driver circuit will be used to demonstrate the basic characteristics of a transistor.

### Material Required

Heathkit Analog Trainer
Voltmeter (VOM)
DC ammeter (may be part of VOM)
1 — Transistor NPN MPSA20 (417-801)
1 — Lamp incandescent (412-601)
1 — Resistor 470 ohm (yellow, violet, brown, gold)
1 — Resistor 1 kilohm (brown, black, red, gold)

## Procedure

1. Turn on the Trainer and adjust the + voltage control for 5 volts. Turn the Trainer off.

2. Construct the circuit shown in Figure 5-28A as shown in Figure 5-28B.

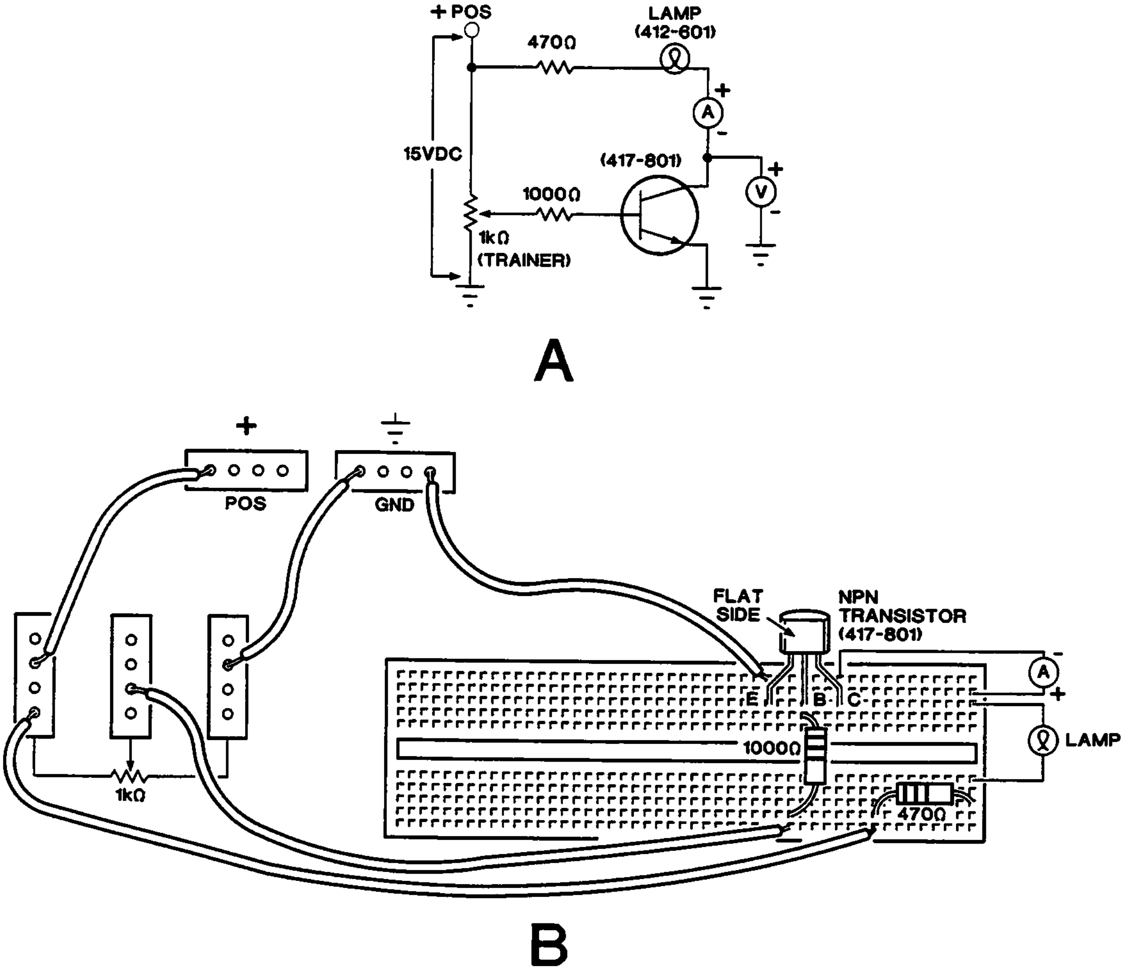


**Figure 5-28**

(A) Lamp driver circuit.

(B) Wiring diagram for lamp driver circuit.

3. Turn on the Trainer. Turn the 1 kilohm potentiometer first in one direction, then in the other. The lamp should light in one direction and turn off in the other, adjust the potentiometer to turn the lamp off. Note: If your circuit does not operate as described, turn off the Trainer and recheck the circuit construction.

4. With the circuit operating properly, make the following voltage and current measurements. Enter the data in Figure 5-29.

   Note: In the next few steps, you will make several voltage adjustments, with he 1 kilohm potentiometer. After each voltage level is set, you will measure the resulting collector current and voltage, and record the data in Figure 5-29. There are two ways in which this sequence can be accomplished. If your voltmeter and ammeter are separate pieces of equipment, you can connect both meters to the locations shown and keep them connected as you adjust the potentiometer and make your measurements. If your voltmeter and ammeter are the same piece of equipment, you will have to switch between two modes of operation. Remember to turn the Trainer off before inserting or removing the ammeter in the circuit.

5. Set the voltmeter range to measure up to 2 Vdc. Refer to Figure 5-28, and connect the meter between the transistors base and ground. Adjust the 1 kilohm potentiometer so that the transistor base voltage ($V_B$) is equal to 0.1 Vdc.

6. Disconnect the voltmeter from the transistor base. Set the meter to measure up to +10 Vdc and connect it to the collector as shown in Figure 5-28. Record the voltage in Figure 5-29, next to the corresponding $V_B$. In this case, 0.1 Vdc.

| $V_B$ (Vdc) | $V_C$ (Vdc) | $I_C$ (mA) | $R_{CE}$ (ohms) |
|---|---|---|---|
| 0.100 | | | |
| 0.300 | | | |
| 0.500 | | | |
| 0.600 | | | |
| 0.650 | | | |
| 0.675 | | | |
| 0.700 | | | |
| 0.725 | | | |
| 0.750 | | | |
| 0.775 | | | |
| 0.800 | | | |
| 0.825 | | | |
| 0.850 | | | |
| 0.875 | | | |

**Figure 5-29**

Chart for recording transistor data.

7. At this time, measure the current $I_C$. If your voltmeter and ammeter are the same piece of equipment, you will have to switch modes of operation. Remember to turn off the Trainer before you connect the ammeter as shown in Figure 5-28. Record your current reading in Figure 5-29, next to the corresponding $V_B$ and $V_C$ readings.

8. Continue to measure and record the voltage and current readings corresponding to different levels of base voltage ($V_B$) until you have completed the Table 1 $V_C$ and $I_C$ columns.

9. Using the values of $V_C$ and $I_C$, as recorded in Figure 5-29, calculate the transistor's collector-emitter resistance ($R_{CE}$) for each $V_B$ setting ($R_{CE}$- $V_C/I_C$). Complete the $R_{CE}$ column of Figure 5-29.

10. Using the data in Figure 5-29, plot the transistor's operating characteristics in Figure 30. First, plot $V_B$ verses $I_C$. Then, plot $V_B$ verses $R_{CE}$.

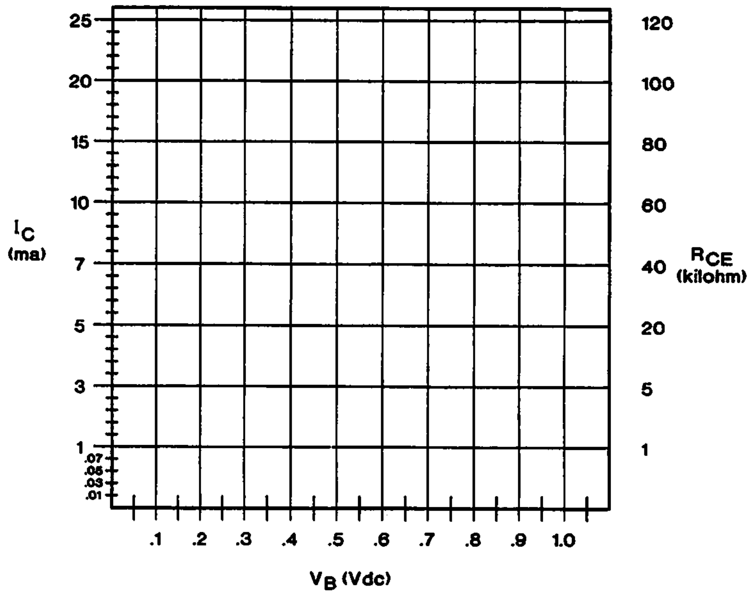


**Figure 5-30**
Transistor Characteristics.

## Discussion

In this experiment, you constructed a simple common-emitter circuit involving an NPN, bipolar transistor. The bipolar transistor is considered to be a current controlled device. This is shown by the small base current produced by base voltage $V_B$. The base current controls the much larger collector-emitter current $I_{CE}$. You can see how the transistor regulates $I_{CE}$ like a variable resistance. In fact, the term transistor comes from trans-(or transfer) resistor. Refer to your graph. The base current ($I_B$) is increased as a result of increasing $V_B$. When $I_B$ is increased, the effective resistance of the transistor from collector to emitter ($R_{CE}$) is decreased. This allows the collector current to increase. When the lamp is visibly lit, the transistor resistance ($R_{CE}$) has been significantly decreased by increasing $I_B$.

From these results, you can recognize how the transistor could be used in switching applications. A small change in base current can be used to switch a large amount of collector current on and off.

# UNIT SUMMARY

Bipolar transistors are two junction devices that consist of three alternately doped semiconductor regions. A transistor that is comprised of a P-type region between two N-type regions is called an NPN transistor. Similarly, a transistor having an N-type region between two P-type regions is a PNP transistor.

The center region of a transistor is called the base and the two outer regions are called the emitter and the collector. The base region is usually much thinner than the other two regions and is lightly doped. The base is P-type material in an NPN transistor and N-type material in a PNP transistor.

Several methods are used in the construction of transistors. The three most commonly used methods are the alloyed method, the diffusion method, and epitaxial growth. Many transistors are constructed simultaneously on a single semiconductor wafer. The wafer is then broken up into individual components. Each component is enclosed in a package to protect the device from mechanical damage and the environment, as well as making it possible to connect the transistor to other circuit components. Transistors are enclosed in many different types of packages, depending on the intended application.

The junction formed between the transistor's emitter and base is often called the emitter-base or emitter junction. Similarly, the junction between the collector and base is called the collector-base or collector junction. Under normal operating conditions, the emitter junction is forward biased and the collector junction is reverse biased. This means that, in an NPN transistor, the N-type emitter must be negative with respect to the P-type base, while the base is negative with respect to the N-type collector.

In a properly biased NPN transistor, electrons flow from the transistor emitter to the base. A few of these electrons leave the base as base current. However, most of the electrons pass through the base to the positive collector as collector current. Therefore, the emitter current is equal to the sum of the base and collector currents. However, when transistor circuits are being analyzed, emitter current and collector current can be considered equal because base current is very low.

Operation of a PNP transistor is similar to that of an NPN transistor. However, the operation of a PNP transistor depends on hole movement instead of electrons. Holes are confined to the internal structure of the transistor. Emitter, base, and collector currents are still produced by electrons flowing into and out of the device. A PNP transistor is properly biased when its emitter is positive with respect to the base and the collector is negative with respect to the base.

Practically all transistor applications are based on a transistor's ability to amplify weak electronic signals. Amplification is the ability to increase the amplitude of an electronic signal. The amplitude of the input signal can be expressed in terms of voltage or current. However, before the transistor can amplify, it must be connected to an external circuit in a manner that allows the input signal to control transistor conduction. The transistor then controls the current through an external load, such as a resistor. A transistor amplifier circuit may be any one of three basic circuit configurations, common emitter, common base, or common collector. In a given configuration, one of the transistor leads, emitter, base, or collector, is the circuit reference point while the other two leads serve as input and output connections.

In a common emitter circuit, the emitter of the transistor is the common reference point while the input is applied to the base and the output is taken from the collector. The common emitter circuit amplifies voltage, current, and power. This circuit configuration has low input resistance and high output resistance.

When a transistor is connected in the common base configuration, the base of the transistor serves as the common reference point. The input voltage is applied to the emitter and the output appears across the collector load resistance. A common base circuit amplifies voltage and power. However, output current is slightly less than input current because collector current is always less than emitter current.

The collector of the transistor serves as the reference point in a common collector circuit. The input is applied to the base and the output is taken from a load resistance connected to the emitter. Common collector circuits amplify current and power. However, the output voltage is always less than the input voltage. This is because part of the input is dropped across the emitter-base junction of the transistor. Common collector circuits cannot be used in applications where voltage amplification is required. However, the high input resistance and low output resistance of this configuration makes the common collector circuit useful in applications where impedance matching is required. The common collector circuit is frequently called an *emitter follower* because the emitter output voltage is in phase, or follows, the input voltage.

Transistors can be easily tested for opens or shorts with an ohmmeter. The ohmmeter should indicate that the forward resistance of each transistor junction is low and that the reverse resistance is high. If a given junction has a high resistance in both directions, the junction is probably open. Conversely, if the ohmmeter indicates that the junction resistance is low in both directions, the junction is probably shorted.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. Bipolar transistors may be classified as

   A. NPN and PIN devices.
   B. NPN and PNP devices.
   C. NNP and PPN devices.
   D. N-type and P-type devices.

2. Bipolar transistors have

   A. two PN junctions.
   B. one PN junction.
   C. three PN junctions.
   D. no PN junctions.

3. A bipolar transistor is constructed so that its base region is

   A. very thick and heavily doped.
   B. very thin and heavily doped.
   C. very thick and lightly doped.
   D. very thin and lightly doped.

4. A transistor's emitter-base junction must always be

   A. subjected to a high reverse current.
   B. biased in the same manner as the collector junction.
   C. forward-biased.
   D. reverse-biased.

5. A transistor's collector-base junction must always be

   A. subjected to a high forward current.
   B. reverse-biased.
   C. forward-biased.
   D. subjected to a low forward current.

6. A transistor's collector current is always

   A. much lower than its emitter current.
   B. lower than its base current.
   C. slightly lower than its emitter current.
   D. higher than its emitter current.

7. When a transistor is connected in a common-base arrangement it is capable of providing

    A. current amplification
    B. current and voltage amplification.
    C. current and power amplification.
    D. voltage and power amplification.

8. A common-emitter circuit provides

    A. current, voltage, and power amplification.
    B. current and voltage amplification only.
    C. voltage and power amplification only.
    D. current and power amplification only.

9. A common-collector circuit may be used to

    A. provide voltage amplification.
    B. provide current and voltage amplification.
    C. match a high impedance to a much lower impedance.
    D. match a low impedance to a high impedance.

10. When checking a good transistor with an ohmmeter, the device should exhibit a

    A. small difference between forward and reverse resistance across each junction.
    B. large difference between forward and reverse resistance across each junction.
    C. a high resistance in either direction across its emitter junction.
    D. a low resistance in either direction across its collector junction.

11. When a transistor has a very low and almost equal forward and reverse resistance between its emitter and base leads, the device effectively has

    A. a good emitter junction.
    B. a shorted collector junction.
    C. a shorted emitter junction.
    D. an open emitter junction.

12. When a transistor has an infinitely high forward and reverse resistance between its base and collector leads, the device effectively has

    A. a good collector junction.
    B. an open collector junction.
    C. a shorted collector junction.
    D. an open emitter junction.

# EXAMINATION ANSWERS

1. B — One type of transistor uses a P-type semiconductor layer which is sandwiched between N-type layers and is referred to as an NPN device. When the opposite arrangement is used the transistor becomes a PNP device.

2. A — Since the bipolar transistor has three layers, it must have two junctions.

3. D — The base region must be thin and lightly doped so that most of the charge carriers can move through the base region and into the collector region.

4. C — The emitter-base junction is forward-biased so that charge carriers (holes or electrons) are forced toward the base region.

5. B — The collector-base junction is reverse-biased so that the collector region is subjected to a potential which will attract the charge carriers arriving at the base.

6. C — The collector current ($I_C$) is only slightly lower than the emitter current ($I_E$) and the difference between these two currents is the base current ($I_B$).

7. D — The common-base circuit provides voltage amplification but no current amplification. However, the increase in voltage means that an increase in power must also occur.

8. A — The common-emitter circuit is the only arrangement that provides an increase in current, voltage, and power.

9. C — The high input resistance and low output resistance of the common-collector circuit makes it suitable for matching a high impedance source to a low impedance load so that an efficient transfer of power is realized.

10. B — The ohmmeter should be connected so that it will forward bias and reverse bias each junction. The meter should indicate that the forward resistance of each junction is relatively low, but the reverse resistance should be quite high. A large difference between forward and reverse resistances should therefore be observed.

11. C — Normally the emitter junction should have a high forward-to-reverse resistance ratio.

12. B — The collector junction should always have a high forward-to-reverse resistance ratio.

## *Unit 6*

# BIPOLAR TRANSISTOR CHARACTERISTICS

# CONTENTS

Introduction . . . . . . . . . . Page 6-4

Unit Objectives . . . . . . . . . . Page 6-4

Unit Activity Guide . . . . . . . . . . Page 6-5

Characteristics of Common-Base Circuits . . . . . . . . . . Page 6-6

  Collector Characteristic Curves . . . . . . . . . . Page 6-6

  Current Gain . . . . . . . . . . Page 6-10

  Alpha Cutoff Frequency . . . . . . . . . . Page 6-12

  Collector-To-Emitter Leakage Current . . . . . . . . . . Page 6-12

Characteristics of Common-Emitter Circuits . . . . . . . . . . Page 6-16

  Collector Characteristic Curves . . . . . . . . . . Page 6-16

  Current Gain . . . . . . . . . . Page 6-19

  Beta Cutoff Frequency . . . . . . . . . . Page 6-21

  Collector-To-Emitter Leakage Current . . . . . . . . . . Page 6-22

Experiment 7 . . . . . . . . . . Page 6-25

Characteristics of Common-Collector Circuits . . . . . . . . . . Page 6-31

  Input Resistance . . . . . . . . . . Page 6-33

  Output Resistance . . . . . . . . . . Page 6-34

Maximum Transistor Ratings . . . . . . . . . . Page 6-36

  Collector Breakdown Voltage . . . . . . . . . . Page 6-36

  Emitter Breakdown Voltage . . . . . . . . . . Page 6-36

  Maximum Collector and Emitter Currents . . . . . . . . . . Page 6-37

  Maximum Collector Dissipation . . . . . . . . . . Page 6-37

- Temperature Ratings . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-37
- Current Gain-Bandwidth Product . . . . . . . . . . . . . . . . . . . . . . . .Page 6-38

Experiment 7A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-40

Experiment 7B . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-44

Experiment 7C . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-48

Unit Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-52

Unit Examination . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-53

Examination Answers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 6-56

# INTRODUCTION

In your previous unit you learned how the bipolar transistor operates and how the device can be used to provide amplification of electronic signals. In this unit you will examine in greater detail some of the bipolar transistor's important electrical characteristics when it is used in each of the three basic amplifier configurations. As explained previously these three arrangements are referred to as the common-base circuit, the common-emitter circuit, and the common-collector circuit. Examine your unit objectives closely and then follow the instructions provided in your unit activity guide as you proceed through this unit.

# UNIT OBJECTIVES

When you have completed this unit on bipolar transistor characteristics you should be able to:

1. Determine the approximate current gain (alpha) of a common-base transistor by using the transistor's collector characteristic curves.
2. Explain the meaning of the term "alpha cutoff frequency."
3. Explain what $I_{CBO}$ is and why it is important.
4. Determine the current gain (beta) of a common-emitter transistor by using the tansistor's collector characteristic curves.
5. Explain the meaning of the term "beta cutoff frequency."
6. Explain what $I_{CEO}$ is and why it is important.
7. Determine alpha when beta is known and vice-versa.
8. Determine the approximate input and output resistance of a common-collector transistor circuit.

# UNIT ACTIVITY GUIDE

**Completion Time**

☐ Read "Characteristics of Common-Base Circuits." __________

☐ Answer Self Test Review Questions 1-11. __________

☐ Read "Characteristics of Common-Emitter Circuits." __________

☐ Answer Self Test Review Questions 12-23. __________

☐ Perform Experiment 7 on "Common-Emitter Amplifier Characteristics-I." __________

☐ Read "Characteristics of Common-Collector Circuits" and "Maximum Transistor Ratings." __________

☐ Answer Self Test Review Questions 24-33. __________

☐ Perform Experiment 7A on "Common-Base Amplifier Characteristics." __________

☐ Perform Experiment 7B on "Common-Emitter Amplifier Characteristics-II." __________

☐ Perform Experiment 7C on "Common-Collector Amplifier Characteristics." __________

☐ Read the Unit Summary __________

☐ Complete the Unit Examination. __________

☐ Check the Examination Answers. __________

# CHARACTERISTICS OF COMMON-BASE CIRCUITS

As explained in the previous unit, a bipolar transistor may be connected in the common-base configuration to provide voltage and power amplification but not current amplification. When a transistor is used in a common-base circuit, an input signal voltage is inserted between the transistor's emitter and base. The input voltage effectively varies the transistor's emitter current which in turn causes the transistor's base and collector currents to vary in a proportional manner. The transistor's emitter and collector currents serve as input and output currents respectively and since the collector current is slightly lower than the emitter current, the common-base circuit cannot provide current amplification. However, a load resistor can be connected between the transistor's collector and base so that the transistor's collector current will flow through the load and develop an output voltage. The voltage developed across this load resistor will be much higher than the input signal voltage thus allowing the circuit to provide voltage amplification. Due to the higher voltage developed across the load resistance, the power supplied to the load will greatly exceed the input power required to operate the circuit thus allowing the circuit to provide power amplification.

The common-base circuit represents just one of three basic circuit arrangements in which the bipolar transistor can be used to provide amplification. The common-base arrangement is not widely used in electronic equipment but it is necessary to examine certain common-base characteristics to obtain an overall understanding of transistor operation. We will now examine some of these important characteristics.

## Collector Characteristic Curves

When a bipolar transistor is connected in the common-base configuration, the transistor's collector current may be controlled by varying the emitter current flowing through the device. However, the transistor's collector current may also be controlled (to a lesser degree) by varying the reverse bias voltage applied to the collector junction. Furthermore, the transistor's collector current tends to vary in direct proportion to changes in emitter current, but the collector current does not always vary proportionally with changes in the reverse bias voltage. Therefore, in order to effectively show the current-voltage relationships that exist within the transistor, it is necessary to plot the related current and voltage values in a graphical manner. When the related values are plotted on a graph, the curve formed shows how the transistor's collector current, emitter current, and reverse voltage are related. This curve shows the transistor's important electrical characteristics when it is connected in the common-base configuration and is therefore referred to as a "characteristic curve."

To adequately show the transistor's electrical characteristics it is necessary to plot a number of characteristic curves. Each curve is plotted for a specific value of emitter current starting at zero and working up to a certain maximum value. In each case the emitter current is held constant while the reverse bias voltage across the collector junction is varied and the change in collector current is observed. The corresponding values are then plotted on a graph and a group of curves are formed.

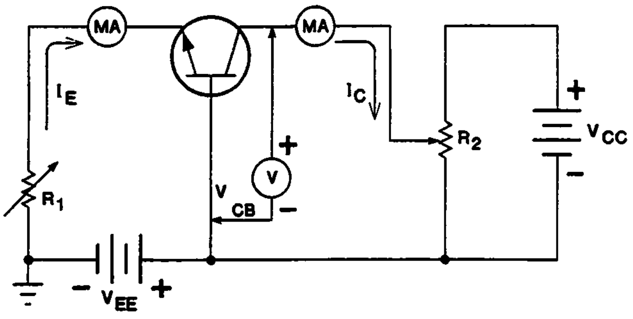


**Figure 6-1**

Circuit used to determine characteristic curves for a common-base transistor

The characteristic curves for a common-base transistor may be obtained by using a circuit like the one shown in Figure 6-1. The transistor used in this circuit is an NPN device; however, the same circuit could be used to plot the curves for a PNP transistor if the polarities of the bias voltages were reversed. Notice that a variable resistor ($R_1$) is in series with the transistor's emitter and the external voltage source ($V_{EE}$) that is used to forward bias the transistor's emitter-base junction. The resistor is used to control the transistor's emitter current ($I_E$). The transistor's collector-base junction is reverse-biased by external voltage source $V_{CC}$ but this voltage is made variable by potentiometer $R_2$. This potentiometer may be adjusted to control the reverse voltage applied to the transistor. Two milliammeters are also used in the circuit to measure the emitter current ($I_E$) and collector current ($I_C$). Also, a voltmeter is used to measure the reverse voltage across the transistor's collector-to-base junction. This reverse voltage is commonly designated as $V_{CB}$ as shown.

A typical set of characteristic curves are shown in Figure 6-2. To obtain these curves $R_1$ is adjusted to various values of emitter current ($I_E$). For each value of emitter current, the transistor's reverse voltage ($V_{CB}$) is varied over a wide range of values and the transistor's corresponding collector current ($I_C$) is observed at each value of $V_{CB}$. When the corresponding $V_{CB}$ and $I_C$ values are graphically plotted for each $I_E$ value, a set of characteristic curves (often called a family of characteristic curves) are obtained. A typical set of these curves are shown in Figure 6-2.

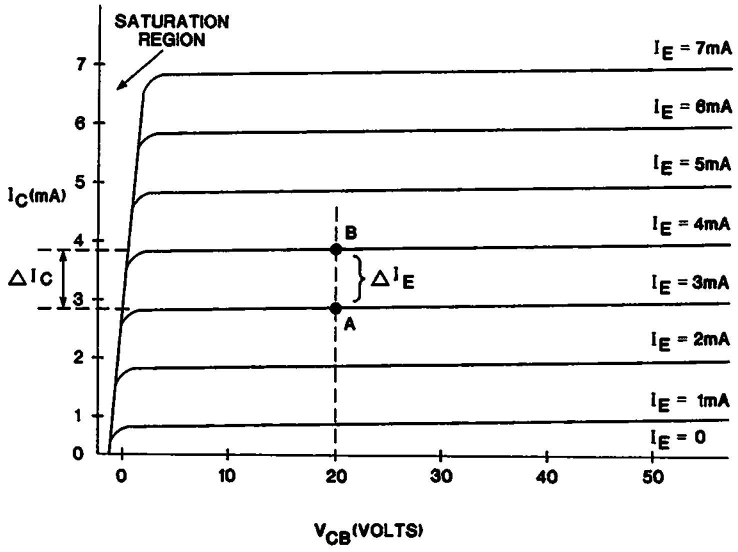


Figure 6-2
Typical collector characteristic curves for a common-base transistor.

Notice that the characteristic curves in Figure 6-2 are plotted for $I_E$ values which range from 0 to 7 milliamperes. The $V_{CB}$ values are plotted horizontally on the graph and the $I_C$ values are plotted vertically. The curve for $I_E = 0$ is produced when no emitter current flows through the transistor. At this time only a small leakage current can flow through the collector-to-base junction. Notice that this current increases only a small amount as $V_{CB}$ is varied from 0 to more than 50 volts and for all practical purposes is nearly equal to $I_E$ (which is equal to zero). Also notice that the collector-to-base junction must be slightly forward biased (as indicated by $V_{CB}$ values to the left of 0 on the horizontal axis of the graph) to completely reduce $I_C$ to zero. This is a peculiar feature of a common-base transistor circuit, and it occurs because of an inherent collector-to-base junction potential which is produced within the transistor.

Notice that the curve for $I_E = 1$ ma rises rapidly and then quickly levels off as $V_{CB}$ changes from a very low forward bias voltage to a reverse bias voltage of more than 50 volts. This indicates that $I_C$ rises to a maximum value and then remains almost constant even though $V_{CB}$ varies over a wide range. Notice that $I_C$ never quite reaches a value of 1 milliampere. This is because $I_C$ can never be exactly equal to $I_E$ in a common-base circuit since a portion of $I_E$ flows out of the transistor's base and becomes base current. The $I_E = 1$ ma curve in Figure 6-2 therefore shows that when $I_E$ is equal to 1 milliampere, $I_C$ will also be approximately 1 milliampere over a wide range of $V_{CB}$ values. Only when $V_{CB}$ is reversed so that it slightly forward biases the transistor's collector-to-base junction, can $I_C$ be reduced to zero.

The remaining curves in Figure 6-2 follow the same general pattern as the $I_E = 1$ ma curve, but at higher values of current. Notice that all of the curves initially overlap but then each curve reaches a maximum value and then remains essentially constant. In each case the maximum current reached (maximum $I_C$ value) is always slightly less than the $I_E$ value. The portion of each curve where collector current rises rapidly with a small change in $V_{CB}$ (before the knee of the curve) is referred to as the "saturation region" as indicated in Figure 6-2.

Operation of the transistor within this region is generally avoided. In most circuits the transistor is biased so that it will operate within the region where $I_C$ is relatively constant while $V_{CB}$ varies (to the right of the knee of each curve). In this region $I_C$ is controlled primarily by $I_E$ (not by $V_{CB}$).

The curves in Figure 6-2 may be used to determine the value of $I_C$ for a wide range of $I_E$ and $V_{CB}$ values. For example, if $V_{CB}$ equals 20 volts and $I_E$ equals 3 milliamperes, then $I_C$ will be approximately equal to 2.8 milliamperes as shown. Under these conditions the transistor is operating at point A on the characteristic curves as indicated. However, the transistor's operating point can be moved to point B as shown by simply increasing $I_E$ to 4 milliamperes while $V_{CB}$ remains at 20 volts. At this time the value of $I_C$ will be approximately 3.8 milliamperes. From these examples you can see how useful the curves are for predicting how the transistor will operate under various conditions.

Manufacturers of transistors will often provide characteristic curves like those in Figure 6-2 to show the transistor's common-base characteristics. However, manufacturers usually test a number of transistors (of a given type) and produce a set of average or typical curves. When the transistor's collector current and voltages are plotted as shown in Figure 6-2, the resulting curves are often referred to as *static collector characteristic curves* or simply *collector characteristic curves.*

The collector characteristic curves shown in Figure 6-2 are for an NPN transistor, but the same type of curves may also be plotted for a PNP transistor. However, in the case of a PNP transistor, the related values ($I_B$, $I_C$, and $V_{CB}$) are generally considered to be negative instead of positive as shown in Figure 6-2.

## Current Gain

The collector characteristic curves just described effectively show the amount of current amplification that a common-base transistor can provide. In other words the collector curves show how much control the transistor's input (emitter) current has over its output (collector) current.

In the previous unit you learned that a common-base circuit cannot actually provide current amplification because its collector current ($I_C$) is always slightly less than its emitter current ($I_E$). However, it is still common practice to describe the common-base transistor in terms of its ability to amplify or provide an increase or gain in current. In other words we may consider the transistor to have a specific current gain but this gain must be slightly less than 1 or unity. This is equivalent to saying that the transistor provides no current gain at all or that it produces a loss in gain.

The current gain of a common-base transistor is determined by varying the transistor's emitter current and observing the corresponding change in collector current. This is done while the transistor's reverse bias voltage ($V_{CB}$) is held constant. The current gain of a common-base transistor is often identified by the Greek letter *alpha* ($\alpha$) and is expressed mathematically as follows:

$$\text{current gain } (\alpha) = \frac{\Delta I_C}{\Delta I_E}$$

This equation simply states that the current gain (alpha) is equal to a small change in collector current ($\Delta I_C$) divided by a small change in emitter current ($\Delta I_E$). It is assumed that the reverse-bias voltage across the collector-to-base junction ($V_{CB}$) remains constant when these measurements are made. Furthermore, the current gain is expressed simply as a ratio of change in the two currents and is not assigned a specific unit of measurement.

The common-base transistor's collector characteristic curves may be used to graphically determine the transistor's current gain. For example, between points A and B on the curves in Figure 6-2, the emitter current ($I_E$) varies from 3 milliamperes to 4 milliamperes (with $V_{CB}$ constant at 20 volts). The collector current ($I_C$) therefore changes from approximately 2.8 milliamperes to approximately 3.8 milliamperes. This means that a change in $I_E$ ($\Delta I_E$) of 4 − 3 or 1 milliampere will produce a change in $I_C$ ($\Delta I_C$) of 3.8 − 2.8 or 1 milliampere. The current gain of the transistor may therefore be expressed mathematically as

$$\text{current gain } (\alpha) = \frac{1 \text{ milliampere}}{1 \text{ milliampere}} = 1$$

In other words the change in $I_E$ is equal to the change in $I_C$ and this results in a gain of 1.

The actual gain of the common-base transistor is not quite equal to 1 as indicated by the previous calculations. Our graphical analysis resulted in a gain of 1 because we could not determine the precise values of $I_C$ from the curves in Figure 6-2. Actually, the change in $I_C$ is slightly less than the change in $I_E$ and the gain of the transistor is just slightly less than 1. Typical values of alpha for most bipolar transistors will range from 0.95 to 0.995.

Manufacturers of transistors usually state the alpha value for each type of transistor that they produce. The alpha value given is usually a minimum expected value which is derived under a specific set of operating conditions (often $V_{CB}$ is 5 volts and $I_C$ is 1 milliampere). Under different operating conditions the alpha value could be slightly different since the transistor's $I_E$, $I_C$, and $V_{CB}$ values do not maintain an exact relationship over the transistor's useful operating range.

The transistor's alpha is also commonly referred to as the *forward current transfer ratio* of the device and is represented by the symbol $h_{fb}$. In fact this designation is becoming increasingly popular while the alpha designation is finding less use.

The alpha ($\alpha$) or forward current transfer ratio ($h_{fb}$) just described is an ac measurement which is derived by observing the corresponding changes in $I_E$ and $I_C$ with $V_{CB}$ constant. However, a similar measurement can be made by using fixed values of $I_E$ and $I_C$. In other words the corresponding values of $I_E$ and $I_C$ (dc or static values) at a specific operating point on the collector curves may be used to calculate the dc alpha value of the transistor. The dc alpha (forward current transfer ratio) of a device is generally represented by the symbol $h_{FB}$.

## Alpha Cutoff Frequency

When a transistor is subjected to a wide range of ac input signals its current gain (alpha) will not remain constant. In general the current gain of the transistor will decrease when the device is subjected to sufficiently high frequencies. When the time for one cycle (period) of the input ac signal approaches the transit time of a charge carrier (the time required for a charge carrier to pass through the device) the gain of the transistor drops rapidly. The frequency at which the gain (alpha) of the transistor drops to 70.7% of its low frequency value is called the *alpha cutoff frequency* and is often represented by the symbol $f_{ab}$. This is an important characteristic of common-base transistor circuits and manufacturers of transistors usually specify the $f_{ab}$ for each type of transistor produced. The $f_{ab}$ value of each transistor is generally determined by measuring the current gain of each device at a reference frequency of 1000 hertz and then the frequency is increased until the gain drops to 70.7% of its reference value.

## Collector-To-Base Leakage Current

The collector characteristic curves in Figure 6-2 show that the value of $I_C$ decreases as $I_E$ decreases. However, when $I_E$ is reduced to zero $I_C$ will not decrease to an absolute zero value. Instead a small leakage current will still continue to flow through the transistor. In fact an extremely small collector current will flow even when the transistor's emitter is open. This action occurs because the transistor's collector-to-base junction acts like a reverse-biased diode and allows a small reverse or leakage current to flow through the junction. As in the case of a reverse-biased diode this leakage current results because of minority carriers in the collector and base regions.

The fact that a small leakage current can flow through a transistor's reverse-biased collector junction even when the transistor's emitter is open is an important consideration. This small leakage current is commonly referred to as the transistor's *collector-to-base leakage current* and is often represented by the symbol $I_{CBO}$. In a transistor $I_{CBO}$ is important because it combines with and therefore becomes part of the transistor's collector current ($I_C$). In other words $I_{CBO}$ still flows through the collector junction even when the transistor is properly biased so that its emitter current ($I_E$) produces a corresponding collector current ($I_C$). The transistor's collector current is therefore made up of two components. It consists of that portion of $I_E$ which does not flow through the transistor's base to become base current ($I_B$) and $I_{CBO}$.

$I_{CBO}$ is affected only slightly by changes in the reverse bias voltage ($V_{CB}$) across the transistor's collector junction. However $I_{CBO}$ is very sensitive to changes in temperature because it is produced by minority carriers. As a general rule $I_{CBO}$ approximately doubles for every 10° centigrade rise in temperature for both silicon and germanium transistors. However, silicon transistors usually exhibit a much lower $I_{CBO}$ value than comparable germanium transistors.

Fortunately $I_{CBO}$ is extremely small in most transistors (often only a few nanoamperes or microamperes). Therefore its effects are minimal in many circuit applications. However in certain critical circuits (especially those sensitive to temperature) its effects must be considered. In our previous discussions and in those which follow the effects of $I_{CBO}$ are considered to be insignificant.

## Self Test Review

1. The collector characteristic curves for a common-base transistor show the relationship between the transistor's collector current ($I_C$) and collector-to-base voltage ($V_{CB}$) for various values of ____________ ____________.
2. The characteristic curves shown in Figure 6-2 are plotted for an NPN transistor.
   a. True
   b. False
3. The characteristic curves in Figure 6-2 show that when $I_E = 0$, $I_C$ will be almost equal to ____________ even though $V_{CB}$ varies over a wide range.
4. The curves in Figure 6-2 show that ____________ is controlled primarily by $I_E$ and is affected only slightly by changes in $V_{CB}$.
5. The portion of each curve where $I_C$ rises rapidly as $V_{CB}$ increases slightly is referred to as the ____________ region.
6. Manufacturers usually test only one transistor of a given type before producing a set of collector characteristic curves for that type of transistor.
   a. True
   b. False
7. The current gain of a common-base transistor is identified by the Greek letter ____________.
8. A transistor's alpha may be determined graphically by comparing a change in $I_E$ ($\Delta I_E$) with a corresponding change in $I_C$ while ____________ remains constant.
9. A transistor's alpha will always be slightly less than ____________.
10. The frequency at which the alpha of a transistor drops to 70.7% of its low frequency value is referred to as the transistor's ____________ ____________ ____________.
11. The small leakage current that flows through a transistor's reverse-biased collector junction when the transistor's emitter is open is represented by the symbol ____________.

## Answers

1. emitter current ($I_E$)
2. (a) True
3. zero
4. $I_C$
5. saturation
6. (b) False — Normally a number of transistors of a given type are tested and a set of average or typical curves is plotted.
7. alpha ($\alpha$)
8. $V_{CB}$
9. 1
10. alpha cutoff frequency
11. $I_{CBO}$

# CHARACTERISTICS OF COMMON-EMITTER CIRCUITS

When a bipolar transistor is connected in a common-emitter configuration, the input signal voltage is inserted between the transistor's base and emitter and effectively varies the transistor's base current which in turn controls the collector current. The base and collector currents serve as input and output currents respectively. Since the collector current is much higher than the base current, the common-emitter circuit provides current amplification. When a load resistor is connected between the transistor's collector and emitter and the collector current is forced to flow through the load, an output voltage is developed across the load which is greater than the input signal voltage. This results in voltage amplification. The circuit also produces a tremendous increase in power because of the increase in current as well as voltage.

The common-emitter circuit is the most widely used transistor circuit configuration in all types of electronic equipment. Since it is an extremely important configuration, a detailed analysis of some of its important electrical characteristics is desirable.

## Collector Characteristic Curves

The various current-voltage relationships in a common-emitter circuit can be quickly analyzed by referring to an applicable set of collector characteristic curves. These curves are plotted in much the same way as the collector curves previously described for the common-base arrangements. However, the curves for the common-emitter circuit show the relationship between the transistor's base current, its collector current, and the voltage across its collector and emitter.

The collector characteristic curves for a common-emitter transistor could be determined by using a circuit like the one shown in Figure 6-3. A potentiometer ($R_1$) is used to adjust the transistor's base current ($I_B$) to various values and at each value of $I_B$ the second potentiometer ($R_3$) is adjusted so that the voltage applied to the transistor's collector and emitter (designated as $V_{CE}$) is varied over a wide range. The resulting change in collector current ($I_C$) is observed in each case. Resistor $R_2$ is used to limit $I_B$ to a safe value. The related currents and voltages are measured with two milliammeters and one voltmeter as shown.

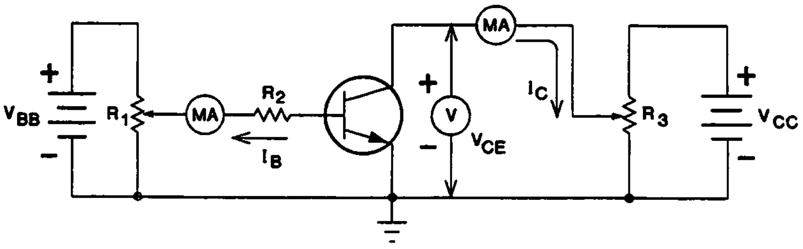


**Figure 6-3**

Circuit used to determine characteristic curves for a common-emitter transistor

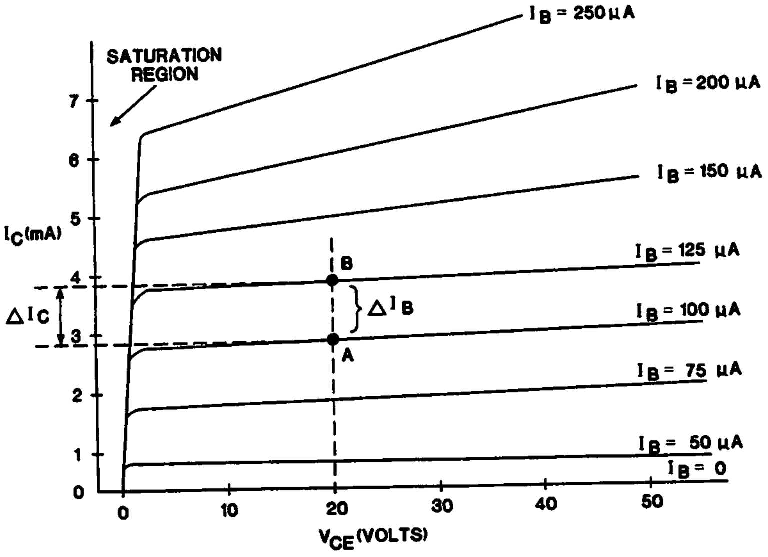


**Figure 6-4**

Typical collector characteristic curves for a common-emitter transistor

A typical set of collector characteristic curves for a common-emitter transistor are shown in Figure 6-4. Notice that these curves are plotted for various $I_B$ values which range from 0 to 250 microamperes. The $V_{CE}$ values are plotted horizontally and the $I_C$ values are plotted vertically. The curves are therefore constructed in the same basic manner as the common-base curves previously described. The curve for $I_B = 0$ is produced when no base current flows through the transistor. At this time only a small leakage current flows through the transistor thus causing $I_C$ to be almost equal to zero. This leakage current increases only slightly as $V_{CE}$ is increased. The curve for $I_B = 50$ microamperes rises rapidly and quickly levels off as $V_{CE}$ is varied from 0 to more than 50 volts. The collector current ($I_C$) reaches a maximum value of approximately 0.7 milliamperes as shown because $I_C$ is limited by the amount of emitter current and base current flowing in the transistor. The emitter and base currents in turn depend on the forward bias voltage applied to the transistor's emitter junction.

The remaining curves plotted in Figure 6-4 are for $I_B$ values that range from 75 to 250 microamperes. Each of these curves have the same general shape as the curve for $I_B = 50$ microamperes. However, the remaining curves have a slightly greater slope. The region just to the left of the knee of each curve is referred to as the *saturation region*. This is the portion of each curve where $I_C$ rises rapidly with an increase in $V_{CE}$. Normally the transistor is biased so that it will operate beyond the knee of each curve or in other words within the region where $I_C$ changes only slightly with changes in $V_{CE}$.

**Figure 6-4A**

Curve Tracer

When used in conjunction with an oscilloscope, this semiconductor curve tracer shown in Figure 6-4A, will automatically plot transistor characteristic curves.

The curves in Figure 6-4 may be used to determine the corresponding values of $I_B$, $I_C$, and $V_{CE}$ at any specific operating point. For example, when $V_{CE}$ is equal to 20 volts and $I_B$ is equal to 100 microamperes, $I_C$ will be equal to approximately 2.5 milliamperes as shown. At this time the transistor is operating at pint A as shown. The $I_C$ value of 2.5 milliamperes is considerably higher than the $I_B$ value of 100 microamperes which is to be expected. This shows that the common-emitter transistor is capable of providing a substantial current gain.

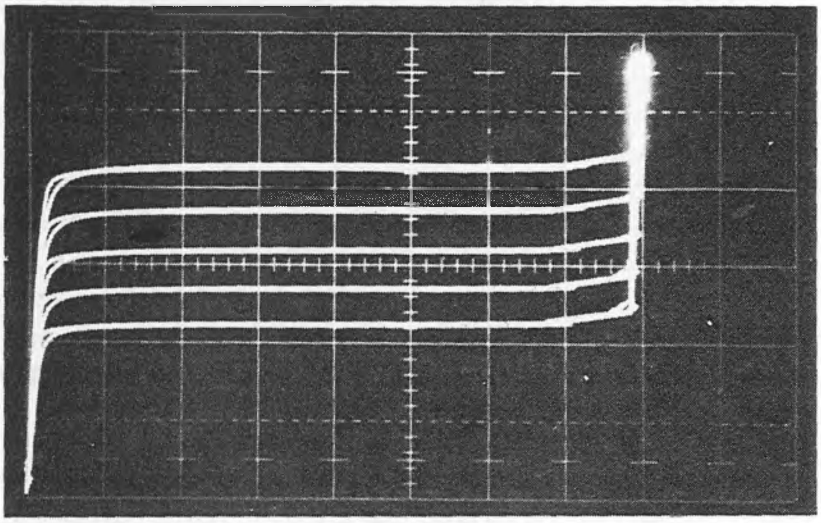


**Figure 6-4B**

An oscilloscope presentation of a set of characteristic curves.

Figure 6-4B shows a set of collector characteristic curves as plotted on an oscilloscope by a semiconductor curve tracer. Here the collector current is plotted as a function of base current and collector-emitter voltage. The sharp upward transition of the collector current shown on the far right represents breakdown in the base-collector junction.

# Current Gain

The amount of current amplification that a common-emitter transistor can provide can be easily determined from the transistor's collector characteristic curves. The transistor's ability to provide an increase or gain in current is determined by varying the transistor's base current and observing the corresponding change in collector current. However, this is done while the transistor's collector-to-emitter voltage ($V_{CE}$) is held constant. The current gain of a common-emitter transistor is sometimes identified by the Greek letter *beta* ($\beta$) and is expressed mathematically as follows:

$$\text{current gain } (\beta) = \frac{\Delta I_C}{\Delta I_B}$$

This equation states that the current gain ($\beta$) is equal to the small change in collector current ($\Delta I_C$) divided by the small change in base current ($\Delta I_B$). However, $V_{CE}$ must remain constant when these corresponding changes in current are measured. The current gain (beta) of a common-emitter transistor is expressed simply as a ratio and is not assigned a specific unit of measurement.

A common-emitter transistor's collector characteristic curves may be used to graphically determine the transistor's current gain (beta). To illustrate this point we will use the curves in Figure 6-4. We will assume that $V_{CE}$ remains constant at 20 volts and that $I_B$ changes from 100 to 125 microamperes. This means that $I_C$ will change from approximately 2.5 milliamperes to approximately 3.6 milliamperes, and the transistor's operating point will move from point A to point B as shown. A total change in $I_B$ of 125-100 or 25 microamperes is therefore accompanied by a total change in $I_C$ of approximately 3.6 − 2.5 or 1.1 milliamperes. When these values are inserted in the equation for beta we obtain a current gain of

$$\text{current gain } (\beta) = \frac{1.1 \text{ milliamperes}}{25 \text{ microamperes}} = 44$$

Our calculations show that the common-emitter transistor represented by the curves in Figure 6-4 has a current gain (beta) of 44. This is a typical beta value for many small and medium power transistors. However, some transistors may have beta values as low as 10 while others may have beta values which exceed 200.

Manufacturers of transistors often specify the beta value for each type of transistor that they produce. The beta values given may be minimum expected values but average and maximum expected values are also sometimes provided. In most cases the manufacturer will also state the transistor's operating conditions along with its beta value since the transistor's beta can change slightly when its operating currents and voltages change.

The beta value just described is a measurement of the transistor's ability to amplify a changing dc or ac current and (like the common-base transistor) is often referred to as the transistor's *forward current transfer ratio.* This ac measurement is also commonly represented by the symbol $h_{fe}$. However, the transistor's beta can also be expressed as a dc measurement by using fixed (dc) values of $I_B$ and $I_C$ instead of changing (ac) values. In this case the corresponding $I_B$ and $I_C$ values are observed at a specific operating point (possibly point A in Figure 6-4) and their ratio is determined. The dc beta (forward current transfer ratio) is often represented by the symbol $h_{FE}$.

When the alpha value of a transistor is known but the beta value is not know, it is possible to determine the beta value mathematically instead of plotting it graphically. The following equation may be used to determine beta when alpha is known.

$$\beta = \frac{\alpha}{1 - \alpha}$$

To illustrate the use of this equation we will assume that a transistor has an alpha of 0.98. When this alpha value is inserted in the equation we find that the transistor has a beta of

$$\beta = \frac{0.98}{1 - 0.98} = 49$$

When a transistor's beta is known but its alpha value is not know, the alpha value may be mathematically determined by using the following equation

$$\alpha = \frac{\beta}{\beta + 1}$$

To illustrate the use of this equation we will assume that a transistor has a beta of 100. When we insert this beta value into the equation we find that the transistor's alpha is equal to

$$\alpha = \frac{100}{100 + 1} = 0.99$$

When a transistor's beta is known and its alpha must be determined, the mathematical method should be used instead of the graphical method previously described. This is because of the difficulty in obtaining accurate readings from the transistor's characteristic curves.

## Beta Cutoff Frequency

Like the common-base transistor previously described, the common-emitter transistor exhibits a decrease in current gain when it is required to amplify ac signals that have sufficiently high frequencies. The frequency required to reduce the common-emitter transistor's current gain (beta) to 70.7% of its low frequency value is referred to as the transistor's *beta cutoff frequency* and is generally represented by the symbol $f_{ae}$. At frequencies above the $f_{ae}$ value, the gain of the transistor is seriously reduced and its performance is therefore degraded.

When any bipolar transistor is alternately connected in a common-base and then a common-emitter arrangement and its respective cutoff frequencies ($f_{ab}$ and $f_{ae}$) are determined, the transistor's $f_{ae}$ will always be much lower than its $f_{ab}$. In other words a common-base circuit is capable of amplifying much higher frequencies than a common-emitter circuit. Unfortunately the common-base transistor's current gain (alpha) is much lower than the common-emitter transistor's current gain (beta).

# Collector-To-Emitter Leakage Current

The collector characteristic curves in Figure 6-4 show that $I_C$ will decrease to a very low value when $I_B$ is reduced to zero. The small leakage current that flows through the transistor at this time is commonly referred to as the transistor's *collector-to-emitter leakage current* and is often represented by the symbol $I_{CEO}$.

Manufacturers often provide the $I_{CEO}$ value of a transistor because it must often be considered when determining how the transistor will operate in certain situations. The $I_{CEO}$ value is determined by using a circuit similar to the one shown in Figure 6-3. However, the base lead of the transistor is left open (disconnected) to insure that $I_B$ is equal to zero.

The leakage current $I_{CEO}$ is comparable to the $I_{CBO}$ value previously described because it is produced by minority carriers and it is temperature sensitive. However, for any given transistor $I_{CEO}$ is much larger than $I_{CBO}$ and it can have a significant influence on the transistor's operation in certain applications. Like $I_{CBO}$, $I_{CEO}$ will combine with $I_C$ and effectively increase the value of $I_C$ when the transistor is biased for normal operation. It is therefore desirable to select transistors with the lowest possible $I_{CEO}$ values to avoid the problems that excessive leakage can produce. A transistor tester as shown in Figure 6-4C, can be used to measure the beta and leakage of a transistor.

**Figure 6-4C**

Transistor Tester.

## Self Test Review

12. The common-emitter transistor's collector characteristic curves show the relationship between the transistor's collector current ($I_C$) and collector-to-emitter voltage ($V_{CE}$) for various values of __________ __________.

13. The collector characteristic curves in Figure 6-4 show that $I_C$ will be slightly less than $I_B$ for any specific value of $V_{CE}$.

    a. True

    b. False

14. The curves in Figure 6-4 show that a small change in $I_B$ can produce a large change in __________ when $V_{CE}$ remains constant.

15. According to the curves in Figure 6-4, $V_{CE}$ has little control over $I_C$ once the transistor is operating beyond the knee of each curve.

    a. True

    b. False

16. Figure 6-4 shows that within the saturation region, $I_C$ increases rapidly as $V_{CE}$ __________ by only a small amount.

17. The current gain of a common-emitter transistor is often represented by the Greek letter __________.

18. The current gain of a common-emitter transistor is expressed as a ratio and has no unit of measurement.

    a. True

    b. False

19. The mathematical expression $\Delta I_C/\Delta I_B$ is used to represent the current gain of the common-emitter transistor.

    a. True

    b. False

20. Typical common-emitter transistors may have current gains that range from __________ to over __________.

21. A mathematical equation may be used to determine a transistor's beta value when its __________ value is known.

22. The frequency required to reduce the common-emitter transistors current gain to 70.7% of its low frequency value is referred to as the transistor's __________ __________ __________.

23. A transistor's collector-to-emitter leakage current is represented by the symbol __________.

## Answers

12. base current ($I_B$)
13. (b) False — $I_C$ will always be much higher than $I_B$ for any value of $V_{CE}$.
14. $I_C$
15. (a) True
16. increases
17. beta ($\beta$)
18. (a) True
19. (a) True
20. 10 to over 200
21. alpha
22. beta cutoff frequency
23. $I_{CEO}$

# EXPERIMENT 7

## Common-Emitter Amplifier Characteristics-I

*OBJECTIVE:*

1. *To demonstrate the operation of a common-emitter transistor.*
2. *To observe and measure some of the important electrical characteristics of a common-emitter transistor.*

### Introduction

The common-emitter configuration is used more often than the common-base and common-collector configurations because it provides current, voltage, and power amplification. An understanding of the basic electrical characteristics of a common-emitter transistor is therefore very important.

In this experiment, you will observe the relationship between a transistor's base current, collector current, and collector-to-emitter voltage. You will begin by taking the necessary measurements and then you will plot the transistor's collector characteristic curves. Then you will use the curves to determine the current gain of the device. Read the entire procedure before performing this experiment.

### Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — Transistor NPN MPSA20 (417-801)
1 — 1 kilohm, resistor (brown, black, red, gold)
1 — 10 kilohm, resistor (brown, black, orange, gold)

## Procedure

1. Plug in your Trainer, be sure that the unit is turned off.

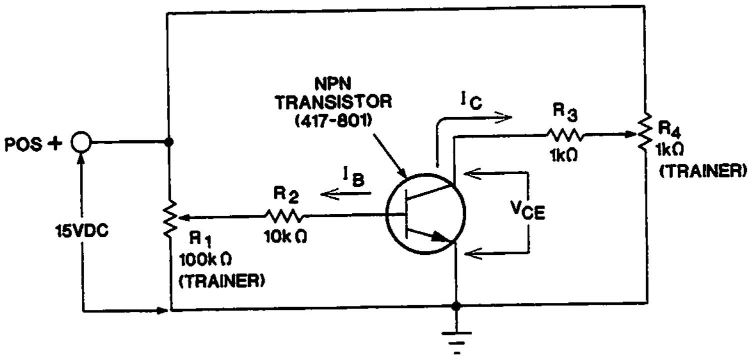


Figure 6-5
Experimental Circuit

2. Construct the circuit shown in Figure 6-5. Use the wiring diagram shown in Figure 6-6. Turn potentiometer $R_1$ for maximum resistance and set $R_4$ to mid-range.

3. Turn the Trainer on and adjust the + voltage control for 15 volts across potentiometer $R_1$. You will be using potentiometer $R_1$ to control the base current ($I_B$) of the NPN transistor and potentiometer $R_4$ to control the transistor's collector-to-emitter voltage ($V_{CE}$). Resistors $R_2$ and $R_3$ are used to limit the currents through the transistor to safe values.

4. You will now adjust the transistor's base current ($I_B$) to a value of 10 microamperes. To do this, you must turn potentiometer $R_1$ slowly until the voltage across $R_2$ (measure this voltage with your VOM) is equal to 0.1 volt. According to Ohm's law, the current through $R_2$ must now equal 0.1 volts divided by 10 kilohms, or 10 microamperes. Since this current also flows through the transistor's base, $I_B$ must equal 10 microamperes at this time. The base current could have been measured with an ammeter, but that would require opening the current path and inserting the ammeter in series with $R_2$ and the transistor's base. Remember, that anytime you add a resistance in series, the current decreases. Since you only have 10 microamperes of current flow in the circuit, it is not a good idea to introduce unnecessary loading into the circuit.

5. Without disturbing the setting of $R_1$ adjust potentiometer $R_4$ until the transistor's emitter-to-collector voltage ($V_{CE}$) is equal to 1 volt.

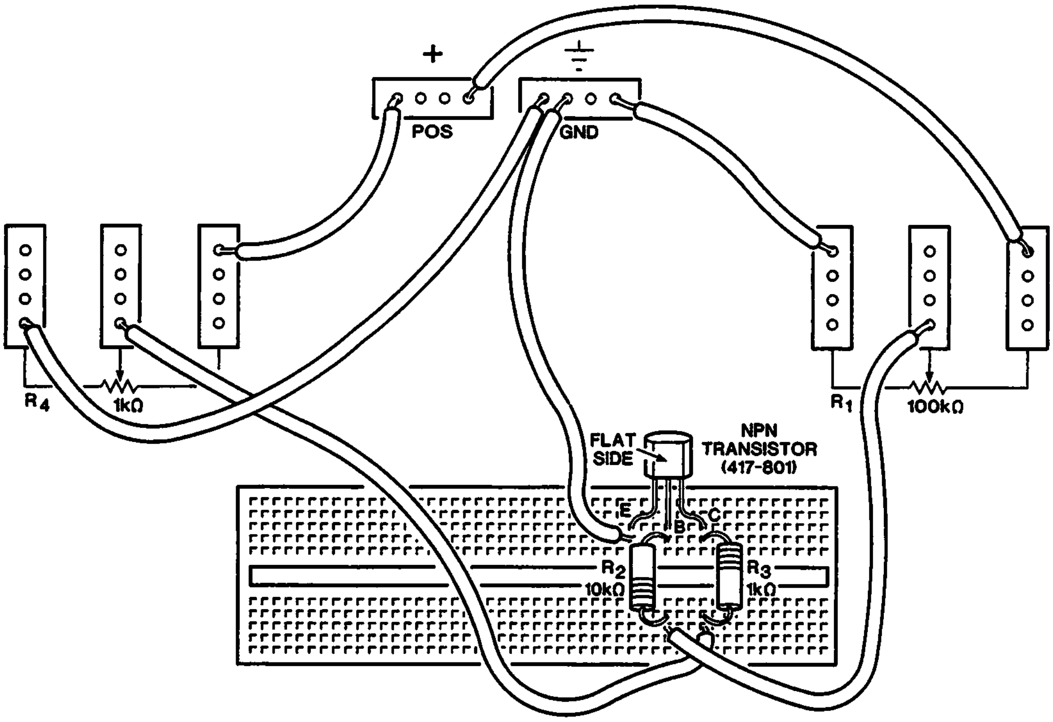


**Figure 6-6**
Wiring diagram for experimental circuit.

6. Now measure the voltage across $R_3$ ($V_{R3}$). Use this voltage measurement and the resistance of $R_3$ (1000 ohms) to calculate the current flowing through $R_3$ ($V_{R3}$ divide by $R_3$). Since $R_3$ is connected to the transistor's collector lead, your calculated value of current also represents the collector current ($I_C$) flowing through the transistor. Record this value of $I_C$ (in milliamperes) immediately below the $V_{CE}$ value of 1 volt in the table provided in Figure 6-7.

| $I_B$ = 10 MICROAMPERES | | | | | | |
|---|---|---|---|---|---|---|
| $V_{CE}$ (VOLTS) | 1 | 2 | 3 | 4 | 5 | 6 |
| $I_C$ (MILLIAMPERES) | | | | | | |

**Figure 6-7**
Table for recording $I_C$ values obtained in steps 4, 5, 6, and 7.

7. Now complete the table in Figure 6-7 by adjusting $V_{CE}$ to the remaining values indicated and recording the corresponding values of $I_C$. To do this simply repeat steps 5 and 6 above for each of the remaining $V_{CE}$ values in the table. When you complete the table in Figure 6-7 you will have a permanent record of $I_C$ values over a range of $V_{CE}$ values when $I_B$ is held constant at 10 microamperes.
8. Now adjust the transistor's base current ($I_B$) to a value of 20 microamperes. To do this simply turn potentiometer $R_1$ slowly clockwise until the voltage across $R_2$ (measure this voltage with your voltmeter) is equal to 0.2 volts. This will cause $I_B$ to have a value of 20 microamperes.
9. Without disturbing the setting of $R_1$, adjust potentiometer $R_4$ until the transistor's emitter-to-collector voltage (measure with your voltmeter) is equal to 1 volt.
10. Now use your voltmeter to measure the voltage across $R_3$. Then use Ohm's law (as you did in step 6) to calculate the current through $R_3$. Your calculated current value represents the transistor's collector current ($I_C$) and should now be recorded (in milliamperes) immediately below the $V_{CE}$ value of 1 volt in the table in Figure 6-8.
11. Complete the table in Figure 6-8 by adjusting $V_{CE}$ to the remaining values indicated and recording the corresponding values of $I_C$. To do this simply repeat steps 9 and 10 for each value of $V_{CE}$ indicated. Your completed table will show how $I_C$ varies over a wide range of $V_{CE}$ values when $I_B$ is equal to 20 microamperes.

| $I_B$ = 20 MICROAMPERES | | | | | | |
|---|---|---|---|---|---|---|
| $V_{CE}$ (VOLTS) | 1 | 2 | 3 | 4 | 5 | 6 |
| $I_C$ (MILLIAMPERES) | | | | | | |

Figure 6-8
Table for recording $I_C$ values obtained in steps 8,9,10 and 11.

12. Now adjust the transistor's base current ($I_B$) to a value of 30 microamperes. To do this turn $R_1$ clockwise until the voltage across $R_2$ (measured with your voltmeter) is equal to 0.3 volts. This will cause $I_B$ to have a value of 30 microamperes.

13. Without disturbing $R_1$, turn $R_4$ clockwise until the transistor's emitter-to-collector voltage (measure with your voltmeter) is equal to 1 volt.

14. Now use your voltmeter to measure the voltage across $R_3$. Then use Ohm's law (as you did in step 6) to calculate the current through $R_3$. This current also represents the collector current ($I_C$) flowing through the transistor. Record this value of $I_C$ (in milliamperes) below the $V_{CE}$ value of 1 volt in the table in Figure 6-9.

Figure 6-9
Table for recording $I_C$ values obtained in steps 12,13,14 and15.

| $I_B$ = 30 MICROAMPERES | | | | | | |
|---|---|---|---|---|---|---|
| $V_{CE}$ (VOLTS) | 1 | 2 | 3 | 4 | 5 | 6 |
| $I_C$ (MILLIAMPERES) | | | | | | |

15. Complete the table in Figure 6-9 by adjusting $V_{CE}$ to the remaining values indicated and recording the corresponding values of $I_C$. To do this simply repeat steps 13 and 14 for each value of $V_{CE}$ indicated. Your completed table will show how $I_C$ varies in relation to $V_{CE}$ when $I_B$ is equal to 30 microamperes.
16. Now use the corresponding values of $I_C$ and $V_{CE}$ that you recorded in Figure 6-7 to plot a collector characteristic curve on the graph shown in Figure 6-10. Simply plot the points where $I_C$ and $V_{CE}$ correspond as indicated in Figure 6-7 and then connect those points to form a continuous curve. Label this curve $I_B$ = 10 microamperes.
17. Use the values of $I_C$ and $V_{CE}$ in Figure 6-8 to plot a second curve on the graph of Figure 6-10. Label this curve $I_B$ = 20 microamperes. Use the values of $I_C$ and $V_{CE}$ in Figure 6-9 to plot another curve in Figure 6-10. Label this curve $I_B$ = 30 microamperes.
18. Now use the set of collector characteristic curves that you just plotted to determine the beta of the transistor. First select a constant value of $V_{CE}$ (above 4 volts). Then note the change that occurs in $I_C$ when $I_B$ changes by a specific amount. Use the equation for beta as shown below and record the beta of the transistor in the space provided below.

$$\text{beta } (\beta) = \frac{\Delta I_C}{\Delta I_B} = \underline{\qquad\qquad}$$

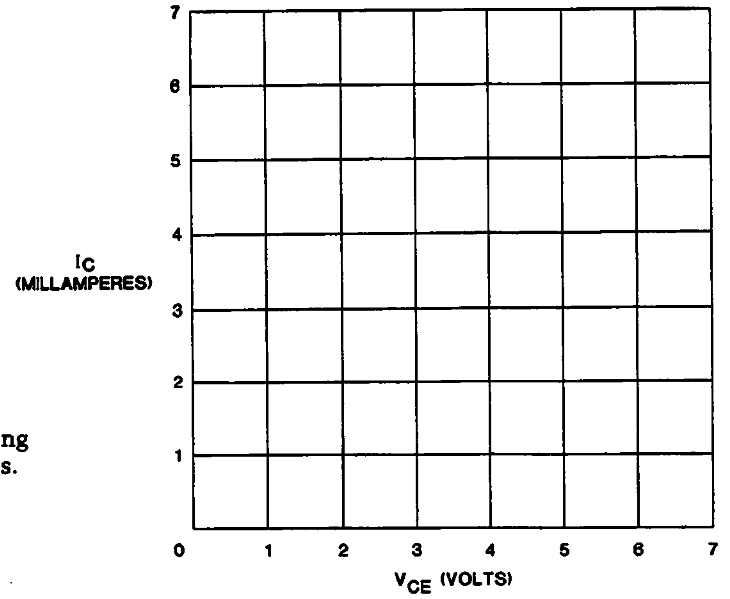


Figure 6-10
Graph for recording $V_{CE}$ and $I_C$ values.

19. Use the beta value that you obtained in the previous step to determine the transistor's alpha value. Simply use the conversion formula given below and record the transistor's alpha value in the space provided below.

$$\text{alpha } (\alpha) = \frac{\beta}{\beta + 1}$$

alpha (α) = ________

20. Your experiment is now complete. Turn off your Experimenter and read the following

## Discussion

In this experiment you first observed the relationship between $I_C$, $I_B$, and $V_{CE}$ in a common-emitter transistor. Your measured specific values of $I_C$ over a range of $V_{CE}$ values for several values of $I_B$ and then plotted these values to obtain a set of collector characteristic curves. The curves that you obtained should have the same general appearance as the ones shown in Figure 6-4 of this unit. However, your curves were not plotted for very low values of $V_{CE}$ (between 0 and 1 volt). Therefore each curve does not show the initial rise in $I_C$ but shows only the relationship between $I_C$ and $V_{CE}$ (for a specific $I_B$ value) after $I_C$ has leveled off (reached the knee of the curve). Each of your curves should therefore show that $I_C$ remains almost constant as $V_{CE}$ is increased. However, in each case the value of $I_C$ should be much higher than the related $I_B$ value for which the curve is plotted.

Next you used your collector characteristic curves to determine the beta of the transistor. You probably obtained a beta that was greater than 100 and possibly as high as 200 thus indicating that your transistor has a very high current gain. The exact value you obtained is not important since beta values can vary over a wide range even for transistors of the same type.

Next you used a conversion formula to calculate the transistor's alpha value. The alpha value that you obtained will be determined by the specific beta value that you derived earlier. However, due to the high gain of the transistor, the alpha value should be greater than 0.99. This means that this transistor would have a current gain that is very close to 1 or unity when it is connected in a common-base configuration.

# CHARACTERISTICS OF COMMON-COLLECTOR CIRCUITS

When a bipolar transistor is connected in a common-collector (also called emitter-follower) arrangement, the input signal voltage is applied between the transistor's base and collector and effectively varies the transistor's base current. The base current in turn controls the transistor's emitter and collector currents but it is the emitter current that is used as the output current. The emitter current flows through a load resistance and the voltage developed across this resistance serves as the output voltage. This output voltage effectively appears between the transistor's emitter and collector. The common-collector arrangement cannot produce an increase in signal voltage because the emitter voltage tends to track or follow the base voltage as explained in a previous unit, however, the circuit arrangement can produce a substantial increase in current and power.

The common-collector arrangement is not used as often as the common-emitter arrangement previously described because it cannot provide voltage amplification. However, the arrangement is used in certain applications where a signal source with a high internal resistance must supply power to a low resistance load. The common-collector circuit with its inherently high input resistance and its low output resistance serves as a buffer and effectively prevents the load from drawing too much current from the source but at the same time allows approximately the same source (signal) voltage to be applied to the load. Since the common-collector circuit allows the same signal voltage supplied by a high resistance source to appear across a low resistance load, the circuit must provide a substantial increase in signal current.

When the common-collector arrangement is compared with the common-emitter circuit, it is possible to mathematically derive an equation which shows how to determine the current gain of the common-collector transistor. We previously discovered that the common-emitter transistor's current gain (beta) is determined by dividing a change in output (collector) current by a corresponding change in input (base) current and is expressed mathematically as

$$\text{current gain } (\beta) = \frac{\Delta I_C}{\Delta I_B}$$

However, the current gain of a common-collector transistor must be equal to a change in output (emitter) current divided by a corresponding change in input (base) current and therefore must be represented mathematically as

$$\text{current gain} = \frac{\Delta I_E}{\Delta I_B}$$

Since a transistor's emitter current must be equal to the sum of its base and collector currents we can express the emitter current as

$$I_E = I_B + I_C$$

Therefore it is possible to express the current gain of a common-collector amplifier as

$$\text{current gain} = \frac{\Delta(I_B + I_C)}{\Delta I_B}$$

which can also be expressed as

$$\text{current gain} = \frac{\Delta I_B + \Delta I_C}{\Delta I_B}$$

This last equation can be further reduced by dividing the $\Delta I_B$ term to obtain the final equation for current gain which is expressed as

$$\text{current gain} = 1 + \frac{\Delta I_C}{\Delta I_B}$$

or

$$\text{current gain} = 1 + \beta$$

Therefore, the current gain of a common-collector transistor is equal to 1 plus the transistor's beta. For all practical purposes when the transistor's beta is high (over 30) the gain of the transistor when it is connected in a common-collector arrangement will be equal to its beta value. Therefore common-collector and common-emitter transistors have approximately the same current gain.

The equation that was just derived is for the ac current gain of the transistor since the transistor's ac beta ($\Delta I_C/\Delta I_B$) is used. However, the transistor's dc beta ($I_C/I_B$) may also be used in the equation to determine the dc current gain of the device used in the common collector configuration.

Manufacturers of transistors seldom provide information relating to all three circuit configurations (common-base, common-emitter, and common-collector). In most cases the manufacturer specifies the features or characteristics of only one circuit arrangement (usually the common-emitter). Therefore it is often necessary to mathematically determine the characteristics of one configuration from information provided on another configuration. The calculations just given show how this is done in the case of current gain. The equation obtained is often referred to as a conversion formula since it shows how to convert from common-emitter gain to common-collector gain. The previous equations which show how to determine alpha when beta is known or beta when alpha is known are also conversion formulas.

## Input Resistance

Although the common-collector circuit exhibits a substantial current gain, this arrangement is most often used because it provides a high input resistance and a low output resistance as previously mentioned. The approximate input resistance of a common-collector circuit can be obtained by using the following equation.

$$R_{in} = \beta R_L$$

This equation states the input resistance ($R_{in}$) is approximately equal to the transistor's beta times the value of the load resistance connected in the transistor's emitter (output) circuit. This load resistance is represented by the designation $R_L$.

A basic common-collector circuit is shown in Figure 6-11. This circuit is drawn in a slightly different manner than the common-collector circuit shown in the previous unit, however it is basically the same circuit. The transistor is biased by voltages $V_{BB}$ and $V_{CC}$ and a load resistance ($R_L$) has been connected between the transistor's emitter and collector. This resistor is effectively between the emitter and collector because $V_{CC}$ is effectively shorted as far as input signals are concerned. The transistor's collector is therefore common to the circuit's input and output. A signal source and its internal resistance ($R_i$) is also connected to the input of the circuit as shown.

We will assume that the transistor in Figure 6-11 has a beta of 50 and that $R_L$ is equal to 2000 ohms. When we substitute these values in the equation given above we obtain an input resistance of approximately

$$R_{in} = (50)\,(2000) = 100{,}000 \text{ ohms}$$

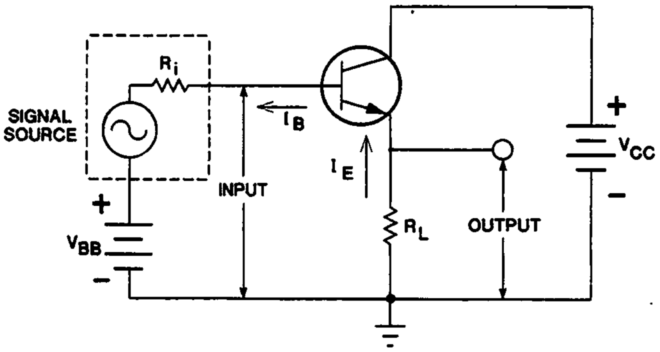


Figure 6-11
A basic common-collector circuit

An input resistance of 100,000 ohms (as determined above) is typical for a common-collector transistor. Although this value may seem high, it can be made much higher by using a transistor with a higher beta or by using a higher value of load resistance.

The value of input resistance (100,000 ohms) just calculated represents only the input resistance of the transistor with the specified value of $R_L$. The input resistance of the circuit will be changed by a substantial amount if external resistors (such as those used to set the transistor's forward bias) were connected in the circuit. In other words, the input resistance of the entire circuit will be different (usually lower) than the input resistance of the transistor.

## Output Resistance

It is much more difficult to calculate the common-collector transistor's output resistance (the resistance seen when looking back into the transistor's output terminals). It is necessary to consider the transistor's gain, its input resistance, and even the internal resistance ($R_i$) of the signal source that is connected to the input of the transistor. Furthermore, this calculation is usually made with $R_L$ removed from the circuit. Then the value of output resistance obtained is considered to be in parallel with $R_L$ and this parallel combination is used to determine the output resistance of the circuit. The two most important factors which control the output resistance of a common-collector circuit are the internal resistance ($R_i$) of the signal source and the transistor's beta. See Figure 6-11. A rough estimate of the output resistance can be obtained by simply dividing $R_i$ by the transistor's beta as shown below.

$$R_{out} = \frac{R_i}{\beta}$$

This equation provides reasonably accurate results as long as $R_i$ is quite large. Once this equation is used to determine $R_{out}$, the parallel resistance of $R_{out}$ and $R_L$ must be calculated to determine the overall output resistance of the circuit.

The value of $R_{out}$ is usually quite low (often less than 100 ohms) in most common-collector circuits. Therefore the overall output resistance of the circuit ($R_{out}$ and $R_L$ in parallel) is generally much lower than the value of $R_L$. Since the simplified equation given above shows $R_{out}$ to be equal to $R_i/\beta$, it follows that the source resistance ($R_i$) is transformed to a lower value by a factor equal to the transistor's beta. However, both the input resistance (described earlier) and the output resistance are affected by changes in signal frequency because the transistor's beta is likewise affected by frequency changes. In general, when the signal frequency exceeds the $f_{ae}$ value, the transistor's beta drops rapidly thus causing $R_{in}$ to decrease and $R_{out}$ to increase. The common-collector arrangement therefore provides a high input resistance and a low output resistance over a frequency range which extends up to $f_{ae}$.

# MAXIMUM TRANSISTOR RATINGS

To this point we have examined some of the important characteristics of each circuit configuration (common-base, common-emitter, and common-collector. However, we assumed that each transistor was working within safe operating limits. Unfortunately, transistor's can be damaged if they are subjected to excessively high currents or voltages and it is important to know just how much current or voltage each device can withstand. Manufacturers usually specify the maximum safe operating currents and voltages for each device and may also specify various power and temperature limitations. These maximum safe values are often referred to as maximum transistor ratings. Some of the most important ratings are described in the following paragraphs.

## Collector Breakdown Voltage

The amount of collector-to-base reverse bias voltage required to produce a sharp increase in collector-to-base current is referred to as the transistor's *collector breakdown voltage*. The sudden increase in current occurs because the transistor's collector junction breaks down in the same way that an ordinary semiconductor diode breaks down when it is subjected to a sufficiently high reverse bias voltage. The collector breakdown voltage is measured with the transistor's emitter open so that $I_E$ is equal to zero and it is usually specified at some value of reverse leakage current. The collector breakdown voltage is often designated as $V_{CBO}$ or $BV_{CBO}$.

## Emitter Breakdown Voltage

The amount of reverse bias voltage required to break down the emitter-to-base junction is referred to as the transistor's *emitter breakdown voltage*. The emitter breakdown voltage is measured with the transistor's collector open so that $I_C$ is equal to zero and it is often represented by the symbol $V_{EBO}$ or the symbol $BV_{EBO}$. Furthermore, $V_{EBO}$ is usually specified at some value of reverse leakage current.

The $V_{EBO}$ value specified by the manufacturer should never be exceeded (even momentarily) or the transistor could possibly be damaged. This same consideration is also true for the $V_{CBO}$ rating previously described. In general, transistors should always be subjected to external voltages which are well below these maximum breakdown ratings.

# Maximum Collector and Emitter Currents

Manufacturers usually specify the maximum value of collector current ($I_C$) and emitter current ($I_E$) that a transistor can safely handle. If these current values are exceeded the transistor could be permanently damaged. The transistor's operating currents should therefore be well below these maximum values.

# Maximum Collector Dissipation

A bipolar transistor dissipates power in the form of heat because it is required to conduct current while being subjected to external voltages. Practically all of the dissipation occurs at the transistor's reverse-biased collector junction and can be easily calculated by multiplying the transistor's collector-emitter voltage by the collector current flowing through the device.

The amount of power that a transistor can safely dissipate at its collector junction is often referred to as the transistor's maximum *collector dissipation rating.* Typical power ratings for transistors will range from several hundred milliwatts to more than 100 watts at an operating temperature of 25° centigrade. When the operating temperature exceeds 25° centigrade, the power rating of the transistor must be derated. In other words it is necessary to calculate the power rating of the device (which will be lower) for the higher operating temperature. This can usually be accomplished by using information supplied by the manufacturer of the device.

# Temperature Ratings

Bipolar transistors are capable of operating within certain temperature ranges. When the temperature of a transistor is too high or too low the device may not operate efficiently and in fact could possibly be damaged. In general, transistors made of silicon materials are capable of operating over a wider temperature range than germanium transistors and it is common practice for manufacturers to specify the safe temperature range of each device. In fact, manufacturers usually indicate the temperature range in which a transistor can be operated or stored.

Manufacturers are usually specific when indicating a transistor's operating temperature range. In most cases the transistor's permissible junction or case temperatures are specified. For example, a typical silicon transistor may have an operating (junction) temperature range of 65° to 200° centigrade. In situations where the transistor's operating temperature tends to approach or exceed the upper end of this range, the transistor is usually attached to a suitable heat sink. The heat sink increases the transistor's ability to radiate heat and therefore helps to cool the device. The temperature affects the power dissipation of the transistor as well as the other characteristics such as beta. Beta is usually proportional to temperature.

## Current Gain — Bandwidth Product

We indicated earlier that the frequency response of a transistor is designated by its alpha or beta cut off frequency. the transistor will provide suitable gain when it is operated below these frequencies. Another way of expressing the frequency response of a transistor is the specification $f_T$, the current gain — bandwidth product. This specification is more often seen on manufacturer's data sheets than the alpha or beta cut off frequencies. The term $f_T$ simply indicates the frequency where the current gain in the common emitter mode is one. This is the maximum operating frequency of the transistor. The current gain — bandwidth product is basically constant, therefore as the operating frequency decreases, the current gain increases by the amount necessary to keep $f_T$ constant.

## Self Test Review

24. The current gain of a common-collector transistor is equal to 1 plus the transistor's________________.

25. If a transistor has a high beta, the current gain of the device will be approximately equal to its beta value when it is connected in a common-collector arrangement.

    a. True
    b. False

26. Manufacturers of transistors usually specify the important characteristics for each of the three basic circuit configurations.

    a. True
    b. False

27. The input resistance of a common-collector transistor is approximately equal to the transistor's beta times the ______________ connected in its emitter circuit.

28. The approximate output resistance of a common-collector transistor may be determined by dividing the __________ ____________ of the signal source connected to the transistor by the transistor's beta.

29. The symbol $BV_{CBO}$ is used to indicate a transistor's ______________ _____________ voltage.

30. The symbol $BV_{EBO}$ is used to indicate a transistor's ______________ _____________ voltage.

31. Most of the power dissipated within a transistor occurs at the transistor's ______________ junction.

32. The amount of power that a transistor can safely dissipate is referred to as the transistor's maximum ______________ ______________ rating.

33. A transistor's operating temperature affects its __________ ___________ rating and ____________.

34. A transistor's upper frequency limit is indicated by ___________.

## Answers

24. beta

25. (a) True

26. (b) False — Manufacturers usually specify the characteristics of only one configuration, common-emitter.

27. load resistance.

28. internal resistance

29. collector breakdown

30. emitter-base breakdown

31. collector

32. collector dissipation

33. power dissipation rating and beta

34. $f_T$

# EXPERIMENT 7A

## Common-Base Amplifier Characteristics

*OBJECTIVE:* *To apply a sine wave signal to a common-base amplifier and observe and measure some of the amplifier's electrical characteristics.*

### Introduction

The common-base configuration is not used as frequently as the other transistor configurations. However, it has characteristics that make it useful in certain applications. Therefore, it is important to understand the characteristics of this circuit configuration.

In this experiment, you will construct a simple common-base audio amplifier circuit and apply a signal to the emitter of the transistor. You will observe both the input and output waveforms and note their phase relationship. You will also calculate the voltage gain of the amplifier circuit.

### Material Required

Heathkit Analog Trainer
Voltmeter (VOM)
Oscilloscope (preferably dual channel)
1 — Transistor NPN MPSA20 (417-801)
1 — 1 kilohm resistor (brown, black, red)
2 — 10 kilohm resistor (brown, black, orange, gold)
1 — 100 kilohm resistor (brown, black, yellow, gold)
1 — 10 microfarad electrolytic capacitor (25-880)

## Procedure

1. Turn on the Trainer and adjust the + voltage control for 15 volts. Turn off the Trainer.

2. Figure 6-12 is the schematic diagram for the circuit in this experiment. Construct the circuit following the wiring diagram in Figure 6-13. In this circuit, $R_1$ and $R_3$ serve as the collector and emitter loads. $R_4$ and $R_2$ set the base voltage or bias for the transistor. $C_1$ couples the signal from the sinewave generator to the emitter of the transistor.

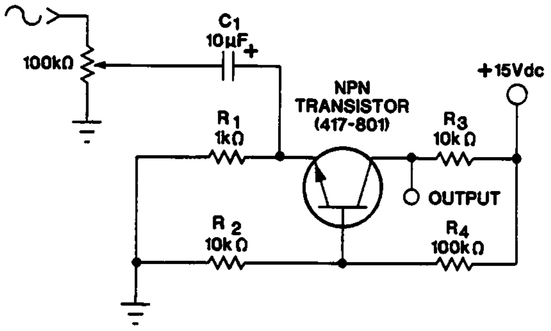


**Figure 6-12**
Experimental Circuit.

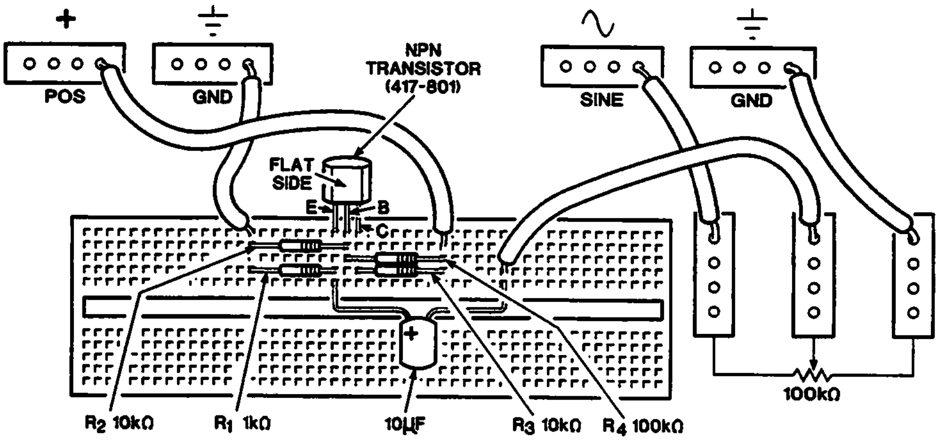


**Figure 6-13**
Wiring diagram for
experimental circuit.

3. Set the following controls on the oscilloscope. The TIME/CM to 1 ms/div. The VOLTS/CM to .1V/div. If you are using a dual trace oscilloscope, set channel 2 for 5V/div. (channel 1 is already set to 0.1 V/div).

4. Set the generator frequency to 200 Hz. Turn on the Trainer.

5. Connect channel 1 of your oscilloscope to the transistor emitter lead and observe the input signal. Adjust the 100 kilohm potentiomenter for a 0.1 Vp-p signal. Adjust the 100 kilohm potentiometer for a 0.1 Vp-p signal. Record the waveform and its amplitude, in volts, in Figure 6-14A.

6. If you are using a dual trace oscilloscope, connect the remaining channel to the collector and view the output signal. Are the two signals in phase? _____________ (yes/no).

   If you are not using a dual trace oscilloscope, change the VOLTS/CM control to 5V/div. Then, move the test lead to the transistor collector and observe the output waveform. Record the signal amplitude and waveform in Figure 6-14B.

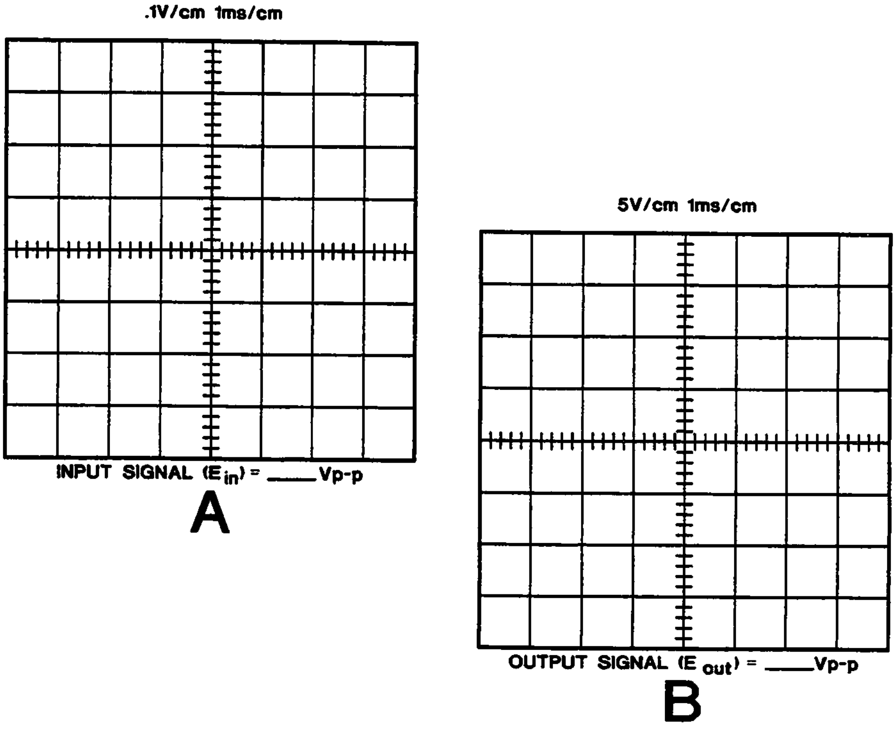

.1V/cm 1ms/cm
INPUT SIGNAL ($E_{in}$) = ____Vp-p
A
5V/cm 1ms/cm
OUTPUT SIGNAL ($E_{out}$) = ____Vp-p
B

**Figure 6-14**
Chart for recording
common base waveforms.

7. Using the value of $E_{in}$ in Figure 6-14A and $E_{out}$ in Figure 6-14B, calculate the voltage gain of the circuit using the formula:

$$\text{Voltage Gain } (A_v) = E_{out}/E_{in}$$

$A_v$ = ________________

Gain is not measured in any specific unit. Gain expresses the ratio of the output signal amplitude to the input signal amplitude.

8. This completes the experiment. Turn off your Trainer and test equipment.

## Discussion

In the experiment you applied a test ac signal to a common-base amplifier circuit. This closely simulates the operation of the common base circuit in a piece of electronic equipment. You observed the input and output waveforms and their amplitude and phase relationships. You noticed that the amplitude of the output signal was larger than that of the input, and that the output and input signals were in phase. In addition, you calculated the voltage gain of the circuit and noted that it is capable of providing voltage amplification.

# EXPERIMENT 7B

## Common-Emitter Amplifier Characteristics-II

*OBJECTIVE:* *To apply a sine wave signal to a common-emitter transistor amplifier and observe and measure some of the amplifiers important characteristics.*

### Introduction

The common-emitter transistor configuration has many characteristics that make it useful for a variety of applications. As a result, it is the most widely used of the three basic transistor configurations. In this experiment, you will construct a common-emitter amplifier circuit and apply an ac signal. You will then subject the circuit to several tests and calculate the voltage gain.

### Material Required

Heathkit Analog Trainer
Voltmeter (VOM)
Oscilloscope (dual channel preferred)
1 — Transistor NPN MPSA200 (417-801)
1 — 100 ohm, resister (brown, black, brown, gold)
2 — 1 kilohm, resistor (brown, black, red, gold)
1 — 10 kilohm, resistor (brown, black, orange, gold)
1 — 10 microfarad electrolytic capacitor (25-880)

# Procedure

1. Turn on the Trainer and adjust the + voltage control for 15 volts. Turn off the Trainer.

2. Figure 6-15 is the schematic diagram for the circuit in this experiment. Construct the circuit following the wiring diagram in Figure 6-16. In this circuit, $R_3$ and $R_4$ are the collector and emitter loads, and $R_1$ and $R_2$ set the bias for the transistor. $C_1$ couples the signal from the sinewave generator to the base of the transistor.

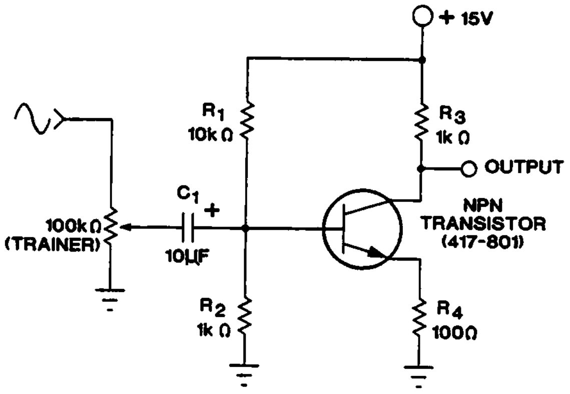


Figure 6-15
Experimental circuit.

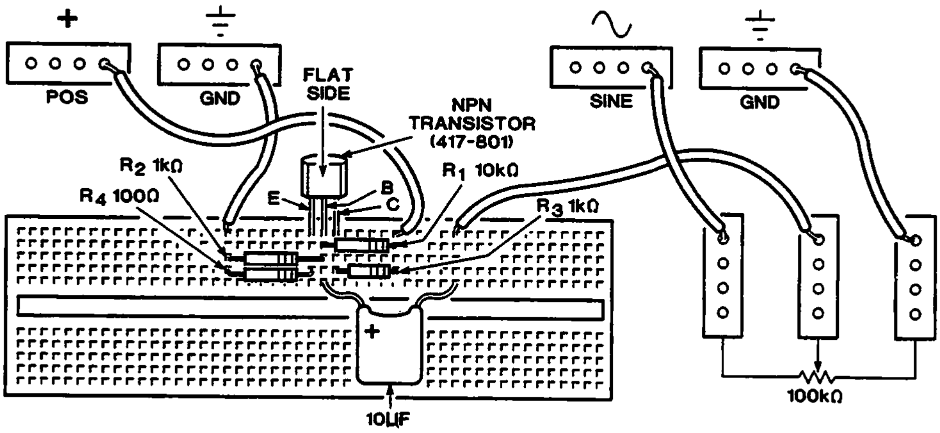


Figure 6-16
Wiring diagram for experimental circuit.

3. Set the following controls on the oscilloscope. The TIME/CM to 1 ms/div. The VOLTS/CM to 1V/div. If you are using a dual trace oscilloscope set channel 2 for 5V/div.

4. Set the generator frequency to 200 Hz. Turn on the Trainer.

5. Connect channel 1 of your oscilloscope to the transistor base lead and observe the input signal. Adjust the 100 kilohm potentiometer for a 1 Vp-p signal. Record the waveform and its amplitude, in volts, in Figure 6-17A.

6. If you are using a dual trace oscilloscope, connect channel 2 to the collector and view the output signal. Are the two signals in phase?) ______________ (yes/no). If you are not using a dual trace oscilloscope change the VOLTS/CM control to 5V/div. Then, move the test lead to the transistor collector and observe the output waveform. Record the signal amplitude and waveform in Figure 6-17B.

7. Using the formula in Experiment 7A and the values of $E_{in}$ and $E_{out}$ in Figure 6-17 calculate the voltage gain of the circuit. $A_V$ = ______________.

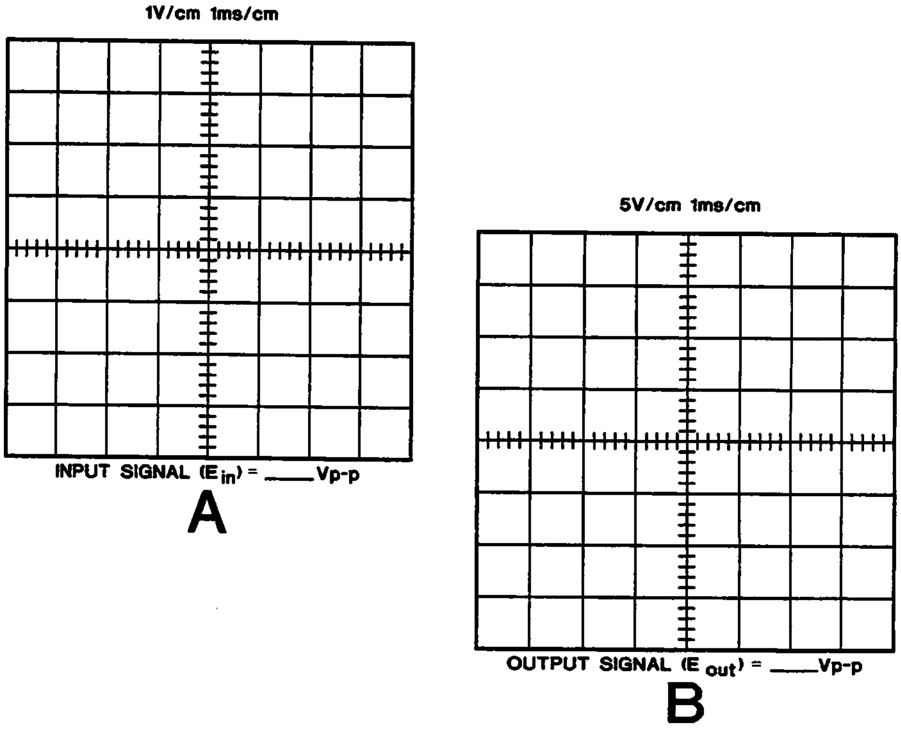

1V/cm 1ms/cm
INPUT SIGNAL ($E_{in}$) = ____ Vp-p
A
5V/cm 1ms/cm
OUTPUT SIGNAL ($E_{out}$) = ____ Vp-p
B

**Figure 6-17**
Chart for recording
common emitter waveforms.

8. This completes the experiment. Turn off your Trainer.

## Discussion

In this experiment, you observed the operation of a common-emitter amplifier and noted the effects of the circuit on the signal. You noticed that the amplitude of the output signal was considerably larger than that of the input signal. A voltage gain of 7-11 is typical. You also noticed that the output signal was 180 degrees out of phase with the input signal. This **phase inversion** of the input signal is the most distinctive characteristic of the common emitter configuration, because the common emitter configuration is the only one of the three basic configurations that inverts the phase of the input signal.

# EXPERIMENT 7C

## Common-Collector Amplifier Characteristics

*OBJECTIVE:* *To apply a sine wave signal to a common-collector amplifier circuit and to observe and measure some of its important characteristics.*

### Introduction

The common-collector circuit has a large number of applications in electronic devices and circuits. Therefore, it is important that you become familiar with its characteristics. Common-collector amplifier circuits are also frequently called emitter-followers.

In this experiment, you will construct an emitter follower circuit on your Trainer and apply a signal to the circuit. You will then observe the input and output signals and note their phase relationship and relative amplitudes. You will also calculate the gain of the circuit.

### Material Required

Heathkit Analog Trainer
Voltmeter (VOM)
Oscilloscope (dual channel preferred)
1 — Transistor NPN MPSA20 (417-801)
1 — 100 ohm resistor (brown, black, brown, gold)
2 — 1 kilohm resistor (brown, black, red, gold)
1 — 10 microfarad electrolytic capacitor (25-880)

## Procedure

1. Turn on the Trainer and adjust the + voltage control for 15 volts. Turn off the Trainer.

2. Figure 6-18 is the schematic diagram for the circuit in this experiment. Construct the circuit following the wiring diagram in Figure 6-19. $R_1$ and $R_2$ develop the bias for the transistor, while $R_3$ serves as the emitter load. $C_1$ couples the sine wave from the signal generator to the base of the transistor.

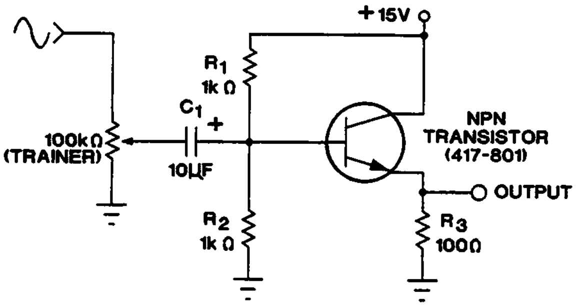


**Figure 6-18**
Experimental Circuit.

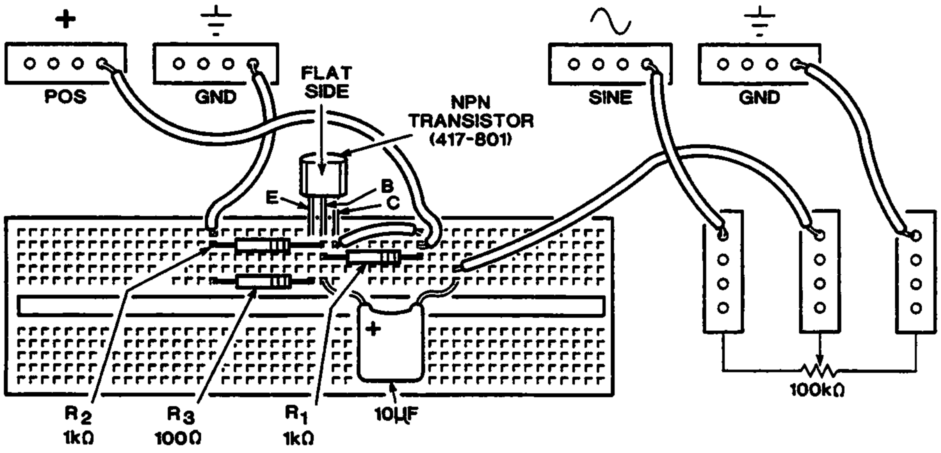


**Figure 6-19**
Wiring diagram for
experimental circuit.

3. Set the following controls on the oscilloscope. The TIME/CM to 1 ms/div. The VOLTS/CM to 1V/div. If you are using a dual trace oscilloscope set channel 2 for 1V/div.

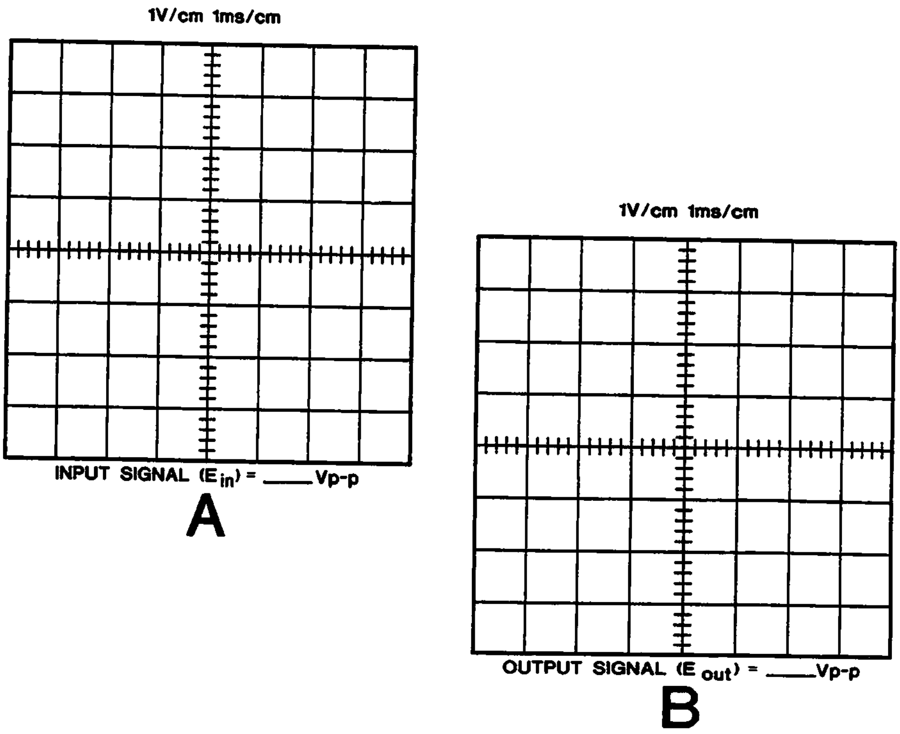
4. Set the generator frequency to 200 Hz. Turn on the Trainer.

5. Connect channel 1 of your oscilloscope to the transistor base lead and observe the input signal. Adjust the 100 kilohm potentiometer for a 1 Vp-p signal. Record the waveform and its amplitude, in volts, in Figure 6-20A.

6. If you are using a dual trace oscilloscope, connect channel 2 to the emitter of the transistor and view the output signal. If you do not have a dual trace scope, move the test lead from the base to the emitter to observe the output waveform. Are the signals in phase? _____________ (yes/no). Do not disturb the scope controls unless it is necessary.

7. Using the values recorded in Figures 6-20A and 6-20B, calculate the voltage gain of the circuit.

$A_V$ = ____________________



INPUT SIGNAL ($E_{in}$) = _____ Vp-p

**A**

OUTPUT SIGNAL ($E_{out}$) = _____ Vp-p

**B**

**Figure 6-20**
Chart for recording
common base waveforms.

8. This completes the experiment 7C. Turn off your Trainer.

## Discussion

In the experiment, you observed the operation of a common-collector amplifier circuit. When you observed the waveforms and calculated the gain, you determined that circuit gain was 1 (unity) or slightly less than unity. This is because the amplitude of the input signal is equal to or slightly greater than that of the output. Even though the common-collector circuit configuration has a voltage gain of less than unity, it is widely used in electronic devices as an impedance matching circuit.

# UNIT SUMMARY

A bipolar transistor may be connected in one of three basic circuit arrangements which are referred to as the common-base circuit, the common-emitter circuit, and the common-collector circuit. The common-base and common-emitter arrangements are used to provide voltage amplification of input signals but the common-collector arrangement is often used to match a high resistance source to a low resistance load.

The amount of current amplification or gain that a transistor can provide can be expressed as a ratio of its input and output currents. In the case of a common-base transistor this ratio is referred to as the transistor's alpha. In the case of a common-emitter transistor, the ratio is known as the transistor's beta. Furthermore, the alpha and beta values for a transistor can be determined by using ac (changing) values of current or dc (fixed) values of current.

A transistor's current gain tapers off when the device is subjected to a sufficiently high signal frequency. The frequency at which a common-base transistor's current gain (alpha) drops to 70.7% of its low frequency value is referred to as the transistor's alpha cutoff frequency. The frequency at which a common-emitter transistor's current gain (beta) drops to 70.7% of its low frequency value is referred to as the transistor's beta cutoff frequency.

Leakage current is undesirable in a transistor but it exists in both NPN and PNP transistors that are made from either silicon or germanium. In general the leakage current is higher in germanium transistors than it is in silicon transistors and in each case the leakage current increases with temperature. The two most important leakage currents that occur in a transistor are represented by the symbols $I_{CBO}$ and $I_{CEO}$. $I_{CBO}$ is the transistor's collector-to-base leakage current when the transistor's emitter is open. $I_{CEO}$ is the transistor's collector-to-emitter leakage current when the transistor's base is open.

The high input resistance and low output resistance of the common-collector transistor makes the device suitable for use in matching resistances. The input resistance of a common-collector transistor is approximately equal to the transistor's beta times the value of the load resistance ($R_L$) connected to the transistor's emitter. The common-collector transistor's output resistance is approximately equal to the internal resistance of the signal source ($R_i$) connected to the transistor's input leads divided by the transistor's beta.

**The most important data on a manufacturer's transistor data sheet is its maximum current, voltage, power, temperature, and frequency ratings.**

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. A common-base transistor's collector characteristic curves are formed by plotting the relationship between $I_C$ and $V_{CB}$ for varius values of

   A. $V_{CE}$
   B. $I_B$
   C. $I_E$
   D. $V_{EB}$

2. A common-base transistor is operating within its saturation region when a small change in $V_{CB}$ produces a large change in

   A. $V_{CE}$
   B. $I_B$
   C. $I_E$
   D. $I_C$

3. A set of collector characteristic curves may be used to graphically determine a common-base transistor's

   A. current gain.
   B. alpha cutoff frequency.
   C. $I_{CEO}$ value.
   D. emitter breakdown voltage.

4. The current gain of a common-base transistor will always be

   A. more than 1.
   B. slightly less than 1.
   C. exactly equal to 1.
   D. less than 0.90.

5. The current gain of a common-base transistor is often referred to as the transistor's

   A. amplification factor.
   B. reverse current transfer ratio.
   C. alpha.
   D. beta.

6. The current gain of a common-base transistor will decrease if the input signal has a

   A. sufficiently low frequency.
   B. sufficiently high frequency.
   C. very low amplitude.
   D. high degree of distortion.

7. In any transistor the quantity $I_{CBO}$ exists because of

   A. majority carriers in the transistor's emitter and base regions.
   B. majority carriers in the transistor's collector and base regions.
   C. minority carriers in the transistor's emitter and base regions.
   D. minority carriers in the transistor's collector and base regions.

8. The collector characteristic curves for a common-emitter transistor show the relationship between the transistor's

   A. $I_C$, $I_B$ and $V_{CE}$ values.
   B. $I_C$, $I_B$ and $V_{CB}$ values.
   C. $I_E$, $I_C$ and $V_{CB}$ values.
   D. $I_E$, $I_B$ and $V_{CE}$ values.

9. The expression $\Delta I_C/\Delta I_B$ is used to determine a common-emitter transistor's

   A. dc alpha
   B. ac alpha
   C. ac beta
   D. dc beta

10. The current gain of a common-emitter transistor

    A. can never exceed 50.
    B. could be as high as 200.
    C. must be less than 1.
    D. is always more than 50.

11. A transistor that has an alpha of 0.97 will have a beta of approximately

    A. 32
    B. 320
    C. 3.2
    D. 64

12. In any given transistor $I_{CEO}$ is likely to be much higher than

    A. any other circuit current.
    B. $I_E$
    C. $I_C$
    D. $I_{CBO}$

13. The input resistance of a common-collector arrangement can be increased by using a

    A. transistor with a lower alpha.
    B. transistor with a lower beta.
    C. higher value of load resistance.
    D. lower value of load resistance.

14. The output resistance of a common-collector arrangement is primarily controlled by

    A. the transistor's beta and the internal resistance of the signal source.
    B. the output load resistance.
    C. the transistor's beta and the output load resistance.
    D. the internal resistance of the signal source and the load resistance.

15. A transistor dissipates power in the form of heat and most of the dissipation

    A. occurs at the emitter junction.
    B. occurs at the collector junction.
    C. is spread evenly throughout the device.
    D. occurs within the emitter region.

16. Another expression for beta in a common emitter transistor circuit is

    A. $h_{fb}$
    B. $f_T$
    C. $h_{FE}$
    D. $I_{CEO}$

# EXAMINATION ANSWERS

1. C — The transistor's collector current ($I_C$) and collector-to-base voltage ($V_{CB}$) values are plotted for different values of emitter current ($I_E$).
2. D — When $I_C$ is below the knee of the curve, the transistor is in its saturation region and any changes in $V_{CB}$ will produce large changes in $I_C$.
3. A — The current gain (alpha) of the common-base transistor may be determined by observing the relationship between $I_C$ and $I_E$ and may be calculated using ac or dc values.
4. B — Common-base transistors will have typical current gains that range from 0.95 to 0.995 or just slightly less than 1.
5. C— The current gain of a common-base circuit is commonly referred to as the transistor's *alpha* but the expression *forward current transfer ratio* is also used.
6. B — The current gain (alpha) will decrease when the period of the input ac signal approaches the transit time of the charge carriers that flow through the device.
7. D — The symbol $I_{CBO}$ represents the transistor's collector-to-base leakage current with the emitter open. This leakage current exists because of the minority carriers in the collector and base.
8. A — The corresponding $I_C$ and $V_{CE}$ values are plotted for various values of $I_B$.
9. C — The quantity $\Delta I_C$ represents a change in $I_C$ and the quantity $\Delta I_B$ represents a change in $I_B$. The ratio of those two changes therefore represents the transistor's ac beta. The transistor's dc beta is calculated by using fixed values of $I_C$ and $I_B$.
10. B — Common-emitter transistors will have typical current gains (betas) that range from 10 to 200.
11. A — Using the conversion formula given in your text the transistor's beta will be equal to

$$\frac{.97}{1 - .97} = 32$$

12. D — The transistor's collector-to-emitter leakage current ($I_{CEO}$) is usually much higher than its collector-to-base leakage current ($I_{CBO}$) and this higher leakage current can have a greater influence on the transistor's operation.

13. C — The input resistance can be increased by either using a transistor with a higher beta or by increasing the load resistance.

14. A — The approximate output resistance is equal to the internal resistance ($R_i$) divided by beta.

15. B — The reverse-biased collector junction dissipates most of the power and essentially determines the transistor's maximum power rating.

16. C — $h_{FE}$ is another designation for beta or the forward current transfer ratio of a transistor used in a common-emitter circuit. You will find the maximum, minimum and typical values of $h_{FE}$ on the manufacturer's data sheet.

## Unit 7

# FIELD EFFECT TRANSISTORS

# CONTENTS

Introduction ........ Page 7-3

Unit Objectives ........ Page 7-3

Unit Activity Guide ........ Page 7-4

The Junction FET ........ Page 7-5

  Construction of a JFET ........ Page 7-5

  Operation of a JFET ........ Page 7-6

  Gate-To-Source Cutoff Voltage ........ Page 7-7

  Pinch-Off Voltage ........ Page 7-8

  Drain Characteristic Curves ........ Page 7-9

    Transconductance ........ Page 7-10

    Symbols ........ Page 7-11

Experiment 8 ........ Page 7-16

The Insulated Gate FET ........ Page 7-22

  Depletion-Mode Devices ........ Page 7-22

  Enhancement-Mode Devices ........ Page 7-26

  CMOS ........ Page 7-29

  Safety Precautions ........ Page 7-31

FET Circuit Arrangements ........ Page 7-35

  Common-Source Circuits ........ Page 7-35

  Common-Gate Circuits ........ Page 7-37

  Common-Drain Circuits ........ Page 7-39

Unit Summary ........ Page 7-41

Unit Examination ........ Page 7-43

Examination Answers ........ Page 7-46

# INTRODUCTION

In this unit, you will examine another important solid-state component that is commonly referred to as a **field effect transistor.** The field effect transistor (commonly called a FET or pronounced fet) operates on a principle that is completely different from that of a conventional bipolar transistor. The FET is a three-terminal device that is capable of providing amplification and it can compete with conventional bipolar transistors in many applications. A basic understanding of FET operation and construction is therefore essential if you are to have a well rounded background in solid-state components.

Basically there are two types of field effect transistors. One type is known as a Junction Field Effect Transistor or simply a JFET. The second type is known as a Metal Oxide Semiconductor Field Effect Transistor (MOSFET). In the following text, The MOSFET is also referred to as a Insulated Gate Field Effect Transistor (IGFET). IGFET was the term used when the device was first developed. You will also be introduced to the Complementary Metal Oxide Semiconductor (CMOS) device. As you proceed through this unit, you will examine the basics of these FET devices. You will learn how they are constructed and how they operate. You will also examine the operation of a typical JFET and CMOS device.

# UNIT OBJECTIVES

When you complete this unit on field effect transistors, you will be able to:

1. Describe how a junction FET operates.
2. Use an FET's drain characteristic curves to determine the transconductance of the device.
3. Properly bias N-channel and P-channel JFETs.
4. Explain the meaning of the expression $V_{GS\,(off)}$ and $V_p$.
5. Explain the difference between the JFET and MOSFET.
6. Describe the difference between a depletion-mode and an enhancement-mode MOSFET (IGFET).
7. Properly bias a depletion-mode and an enhancement-mode MOSFET (IGFET).
8. Define the term CMOS.
9. Identify the schematic symbol of a CMOS device.
10. Name three basic FET circuit arrangements.

# UNIT ACTIVITY GUIDE

| | Completion Time |
|---|---|
| ☐ Read "The Junction FET." | ________ |
| ☐ Answer Self Test Review Questions 1-20. | ________ |
| ☐ Perform Experiment 8 on "Junction FET Characteristics." | ________ |
| ☐ Read "The Insulated Gate FET." | ________ |
| ☐ Answer Self Test Review Questions 21-45. | ________ |
| ☐ Read "FET Circuit Arrangements." | ________ |
| ☐ Answer Self Test Review Questions 46-51. | ________ |
| ☐ Read the Unit Summary. | ________ |
| ☐ Complete the Unit Examination. | ________ |
| ☐ Check the Examination Answers. | ________ |

# THE JUNCTION FET

The junction FET (also called a JFET) finds many applications in electronic circuits. This device is constructed from N-type and P-type semiconductor materials and like the conventional bipolar transistor, it is capable of amplifying electronic signals. However, the junction FET is constructed in a different manner than a bipolar transistor and the device operates on an entirely different principle.

A basic knowledge of junction FET construction is necessary in order to understand how the device operates. Therefore, we will first consider the physical aspects of the junction FET before we consider its electrical characteristics.

## Construction of a JFET

The construction of a junction FET begins with a lightly doped semiconductor material (usually silicon) which is referred to as a *substrate*. The substrate simply serves as a platform on which the remaining electrodes are formed and the substrate can be either a P-type or an N-type material. Through the use of diffusion and epitaxial growth techniques, an oppositely doped region is formed within the substrate to create what is effectively a PN junction. However; it is the unique shape of this PN junction that is important.

The structure that is created by the process described above is shown in Figure 7-1. The region that is embedded in the substrate material is U-shaped so that it is flush with the upper surface of the substrate at only two points.

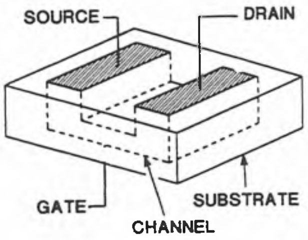


Figure 7-1
Basic construction of a Junction FET.

The embedded region actually forms a channel of oppositely doped semiconductor material through the substrate. Therefore, the embedded region is generally referred to as simply a *channel* as indicated in Figure 7-1. When this channel is made from an N-type material and embedded in a P-type substrate, the entire structure is known as an N-channel junction FET. However; when a P-type channel is implanted in an N-type substrate, the device becomes a P-channel junction FET. Therefore, a junction FET is similar to a conventional bipolar transistor in that it can be constructed in two different ways. The junction FET is either an N-channel or a P-channel device while the bipolar transistor is either an NPN or a PNP device.

The construction of a basic junction FET is completed by making three electrical connections to the device as shown in Figure 7-1. One wire or lead is attached to the substrate as shown. This connection is referred to as the *gate*. Then wires are attached to each end of the channel as shown. These two leads are referred to as the *source* and the *drain*. In most junction FET's the channel is geometrically symmetrical and it makes no difference which end of the channel is used as the source and which is used as the drain. Only in special types of FET's will the channel be asymmetrical and when using these special devices the source and drain leads cannot be interchanged.

## Operation of a JFET

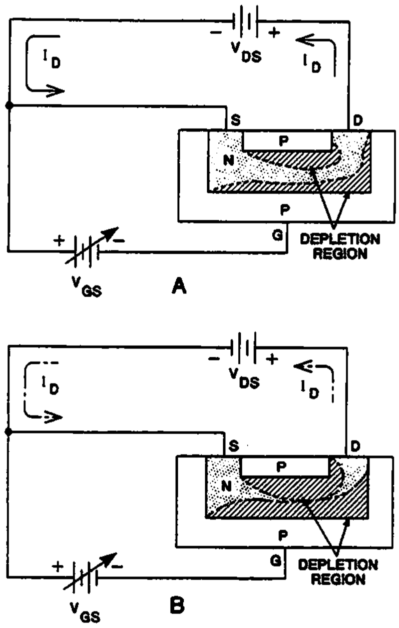


Figure 7-2

A properly biased N-channel Junction FET

A junction FET (JFET) is a *unipolar device*, the charged carriers have only one polarity. On the other hand, a conventional PNP or NPN transistor is bipolar, having both majority and minority charged carriers. Never the less, a JFET requires two external bias voltages for proper operation, just like the conventional transistor. One voltage source is normally connected between the source and drain, so that a current is forced to flow through the channel within the device. The second voltage is applied between the gate and the source and is used to control the amount of current flowing through the channel.

Refer to Figure 7-2A. This figure shows a cross section view of an N-channel junction FET and its required operating voltages. Notice that an external voltage source is connected between the drain (D) and the source (S) leads and this drain-to-source voltage is represented by the symbol $V_{DS}$. The voltage $V_{DS}$ is connected so that the source is made negative with respect to the drain. This voltage causes a current to flow through the N-type channel because of the majority carriers (free electrons) within the N-type material. This source-to-drain current is commonly referred to as the FET's *drain current* and is represented by the symbol $I_D$. The channel simply appears as a resistance to the supply voltage $V_{DS}$.

Also, notice in Figure 7-2A that a voltage is applied between the gate (G) and the source (S) of the FET. This gate-to-source voltage (designated as $V_{GS}$) causes the P-type gate to be negative with respect to the N-type source. Since the source is effectively just one end of the N-type channel, $V_{GS}$ effectively reverse biases the PN junction formed by the P-type gate and the N-type channel. This reverse bias voltage causes a depletion region (an area devoid of majority carriers) to form within the vicinity of the PN junction. As shown in Figure 7-2A, this depletion region spreads inward along the length of the channel. Although two depletion regions appear to exist, only one is created. This depletion region extends around the wall of the N-type channel, since all sides of the channel are in contact with the P-type substrate (which serves as the gate). Furthermore, the depletion region will be somewhat wider at the drain end of the channel than at the source end. This is because $V_{DS}$ effectively adds to $V_{GS}$ so that the voltage across the drain end of the PN junction is higher than the voltage across the source end of the junction.

The size of the depletion region in Figure 7-2A is controlled by voltage $V_{GS}$. When $V_{GS}$ increases, the depletion region increases in size. When $V_{GS}$ decreases, the depletion region decreases in size. Furthermore, when the depletion region increases in size, the N-type channel is effectively reduced in size (fewer free electrons are available) and less current will be able to flow through the channel. The opposite is true when the depletion region decreases in size. This means that $V_{GS}$ may be used to effectively control the drain current ($I_D$) flowing through the channel. An increase in $V_{GS}$ results in a decrease in $I_D$ and vice-versa. We can say that $V_{GS}$ controls the resistance of the channel. Remember that $V_{GS}$ reverse biases the PN junction formed by the gate and the channel. Therefore, only an extremely small (almost insignificant) leakage current flows from the gate to the source.

It is important to note that a voltage ($V_{GS}$) is used to control the drain current ($I_D$) in a junction FET. In normal operation the voltage applied between the gate and the source ($V_{GS}$) serves as an input voltage, to control the device. The drain current ($I_D$) represents the output current which can be made to flow through a load. This action is considerably different from the action that takes place in a bipolar transistor. In the transistor, an input current (not a voltage) is used to control an output current. Also; since the gate-to-source junction of the FET is reverse-biased by $V_{GS}$, the FET has an extremely high input resistance. This is just the opposite of a bipolar transistor which has a forward-biased emitter-to-base junction and therefore a relatively low input resistance. The gate-to-source voltage ($V_{GS}$) must never be reversed so that the PN junction formed by the gate and the channel becomes forward-biased. This would result in a relatively large current through the junction which would cause the input resistance of the device to drop to a low value and the gain of the device would also be significantly reduced.

## Gate-to-Source Cutoff Voltage

As mentioned earlier, when $V_{GS}$ is increased, the depletion region within the FET increases in size and allows less drain current ($I_D$) to flow. If $V_{GS}$ is increased to a sufficiently high value, the depletion region is increased in size until the entire channel is depleted of majority carriers as shown in Figure 7-2B. This causes $I_D$ to decrease to an extremely small value and for all practical purposes is reduced to zero.

The gate-to-source voltage required to reduce $I_D$ to zero (regardless of the value of $V_{DS}$) is referred to as the *gate-to-source cutoff voltage* and is represented by the symbol $V_{GS(off)}$. Manufacturers of FET's usually specify the applicable cutoff voltage for each type of FET that they produce.

## Pinch-Off Voltage

The drain-to-source voltage ($V_{DS}$) also has a certain amount of control over the depletion region within the junction FET. The effect of $V_{DS}$ can be noted in Figures 7-2A and 7-2B since the depletion region is wider near the drain than it is near the source. As explained earlier, this occurs because $V_{DS}$ is in series with $V_{GS}$ thus causing a greater voltage to exist across the PN junction near the drain. If $V_{DS}$ increases in value, the action becomes even more pronounced and can affect the drain current ($I_D$) flowing through the device.

If $V_{DS}$ is increased from zero to higher values, $I_D$ is also increased. However, a continued increase in $V_{DS}$ will not result in a constant rise in $I_D$. Instead, a point is soon reached where $I_D$ levels off and then increases only slightly as $V_{DS}$ continues to rise. This action occurs because the size of the depletion region increases (especially near the drain) as $V_{DS}$ increases and eventually a point is reached where the N-type channel becomes so depleted of majority carriers that it will not allow $I_D$ to increase proportionally with $V_{DS}$. In other words a point is reached where the resistance of the channel effectively begins to increase as $V_{DS}$ increases thus causing $I_D$ to increase at a much slower rate.

Since $I_D$ levels off because the depletion region expands and effectively reduces the channel width, $I_D$ is said to be *pinched-off*. The value of $V_{DS}$ required to pinch off or limit $I_D$ is referred to as the *pinch-off voltage* and is represented by the symbol $V_P$. Manufacturers usually provide the value of $V_P$ for a given FET for a gate-to-source voltage ($V_{GS}$) of zero. $V_P$ is measured by shorting the gate and source leads. This means that $I_D$ will increase up to its maximum possible value (when, $V_{GS} = 0$) and then level off. When $V_{GS}$ is equal to zero, the drain current flowing through the FET is often identified as $I_{DSS}$ instead of $I_D$. Furthermore, manufacturers often provide the FET's $I_{DSS}$ value when $V_{DS}$ is equal to or greater than $V_P$. In this case $I_{DSS}$ represents the FET's maximum drain current when $V_{GS}$ is equal to zero.

In practice, the value of $V_P$ (when $V_{GS} = 0$) will always be close to the value of $V_{GS(off)}$ for any given FET. In fact these two quantities may be interchanged in any calculation in which either quantity is involved. This means that when $V_{GS}$ is equal to or greater than $V_P$; the drain current ($I_D$) will be effectively reduced to zero. Also, when $V_P$ is equal to the $V_{GS(off)}$ value, the drain-current ($I_D$) flowing through the device will be effectively pinched off.

## Drain Characteristic Curves

In the previous unit you saw how a set of collector characteristic curves could be used to show the relationship between the input and output currents and the output voltage associated with a bipolar transistor. A similar set of curves can be used to show the relationship between $V_{GS}$, $V_{DS}$, and $I_D$ in a junction FET. In this case they referred to as drain characteristic curves.

A typical set (sometimes called a family) of drain characteristic curves is shown in Figure 7-3. These curves show how $I_D$ and $V_{DS}$ vary in relation to each other for various values of $V_{GS}$. Each curve is formed by alternately adjusting $V_{GS}$ to a specific value and then increasing $V_{DS}$ from zero to some maximum value while observing the change in $I_D$. Notice that when $V_{GS}$ is equal to zero, $I_D$ increases rapidly as $V_{DS}$ increases from zero. However, $I_D$ soon levels off at some maximum value as shown. When this point is reached, the corresponding $I_{DSS}$ and $V_P$ values previously described are obtained.

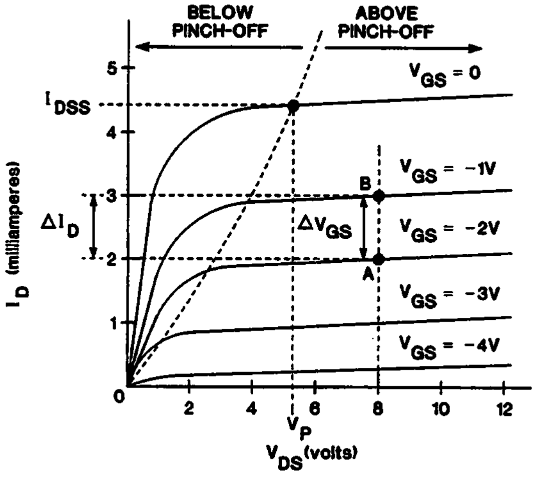


Figure 7-3
Typical drain characteristic curves for an N-channel FET

The remaining curves in Figures 7-3 are plotted for higher values of $V_{GS}$. Notice that for each higher value of $V_{GS}$, $I_D$ levels off at a lower value so that the corresponding pinch-off voltage ($V_P$) must also be lower. The dashed line that curves upward and to the right crosses each curve at the approximate pint where $I_D$ is pinched off. When the FET is biased so that it is operating to the left of this dashed line, the device is said to be operating *below pinch-off*. This region to the left of the dashed line is sometimes referred to as the *ohmic region* or the *triode region*. When the

FET is biased to operate to the right of the dashed line, the device is operating *above pinch-off*. This region to the right of the line is often referred to as the *pinch-off region*. In most applications the FET is biased so that it operates above pinch-off or in other words within the pinch-off region. Operation within the pinch-off region is assured by simply making $V_{DS}$ higher than $V_P$ or $V_{GS(off)}$.

As shown in Figure 7-3, the drain current ($I_D$) will be maximum (for any specific value of $V_{DS}$) when $V_{GS}$ is equal to zero. However, $I_D$ will decrease as $V_{GS}$ increases in value. As explained earlier, this action results because the channel within the FET becomes depleted of majority carriers as $V_{GS}$ increases, thus offering a higher resistance to the flow of current. Since the operation of the junction FET is controlled by varying the depletion region within the device, the junction FET is said to operate in the *depletion mode*.

## Transconductance

**Like the bipolar transistors that were discussed previously, the FET is most often used to amplify electronic signals. In the case of a bipolar transistor, this amplifying ability can be expressed mathematically as a ratio of input and output currents (the transistor's alpha or beta). In the case of an FET a similar mathematical relationship can also be used.**

The amplifying ability of an FET is measured by noting the effect that the gate-to-source voltage ($V_{GS}$) has on drain current ($I_D$). $V_{GS}$ is varied a small amount and the corresponding change in $I_D$ is observed. Then these two quantities are expressed as a mathematical ratio. This ratio is commonly referred to as the FET's *transconductance* and is expressed mathematically as

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

This equation simply states that the transconductance (designated as $g_m$) is equal to a small change in $I_D$ divided by a corresponding change in $V_{GS}$. Although not shown in this equation, the FET's drain-to-source voltage ($V_{DS}$) must be held constant when these changes are observed. Furthermore, the transconductance is expressed in units called *mhos*.

Although the quantities $\Delta I_D$ and $\Delta V_{GS}$ could be measured in a test circuit, these quantities can also be determined graphically by referring to an applicable set of drain characteristic curves. To demonstrate how this is accomplished we will use the typical curves shown in Figure 7-3. We will determine the transconductance of the FET within the pinch-off region (above pinch-off) since this is the region most commonly used.

We will assume that $V_{DS}$ remains constant at 8 volts and that $V_{GS}$ changes from –1 volt to –2 volts as indicated at points A and B in Figure 7-3. This change in $V_{GS}$ causes $I_D$ to change from approximately 3 to 2 milliamperes. A total change in $V_{GS}$ of 2-1 or 1 volt therefore corresponds to a change in $1_D$ of 3-2 or 1 milliampere. When these corresponding changes in $I_D$ and $V_{GS}$ are inserted in the equation given above, we obtain a transconductance of:

$$g_m = \frac{1 \text{ milliampere}}{1 \text{ volt}}$$

$$g_m = \frac{.001}{1}$$

$$g_m = .001 \text{ mho or } 1000 \text{ micromhos}$$

The transconductance is therefore equal to 1000 micromhos between points A and B. however. this value will vary slightly at different operating points on the curves. This is because the curves are not equally spaced within the pinch-off region.

The greater the change in $I_D$ for a change in $V_{GS}$, the higher the gain of the FET. In general a high $g_m$ is a desirable FET characteristic.

## Symbols

To this point we have examined only the N-channel junction FET. The operation of a P-channel FET has not been discussed because it operates in the same manner as the N-channel device and it has the same basic characteristics. The primary difference is in the manner in which the drain current ($I_D$) flows through the channel. In a P-channel FET, $I_D$ is supported by the movement of holes in a P-type channel. However, these holes are still the majority carriers within the P-type channel just as the electrons are the majority carriers in the N-channel FET. Also, the P-channel FET has a gate (or substrate) that is formed from N-type material. This of course is opposite to the conditions that exist within an N-channel FET. This means that the polarities of the bias voltages ($V_{GS}$ and $V_{DS}$) are exactly opposite for the N-channel and P-channel devices.

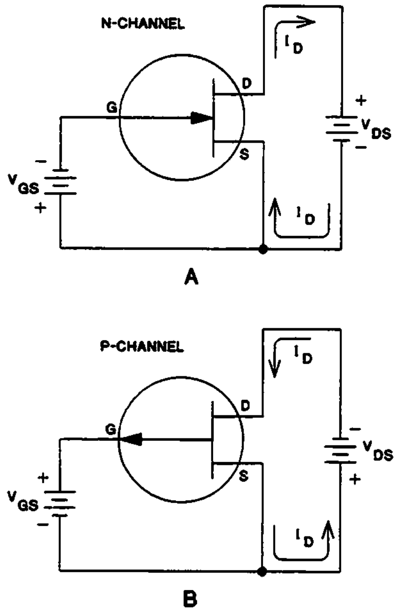


Figure 7-4
Schematic representation of properly biased N-channel (A) and P-channel (B) junction FET's

An N-channel junction FET and a P-channel junction FET are shown in Figure 7-4 as they would appear in a circuit diagram or schematic. The required bias voltages for each device are also shown. Notice that the symbol used for each device is almost identical. The only difference is the direction of the arrow on the gate (G) lead. The N-channel junction FET symbol shown in Figure 7-4A uses an arrow that points inward. However, the P-channel junction FET symbol in Figure 7-4B uses an arrow that points outward. Also notice that the polarities of the bias voltage are exactly opposite. The N-channel FET must be biased so that its drain (D) is positive with respect to its source (S) and its gate (G) must be negative with respect to the source. This negative potential on the gate accounts for the use of the minus signs before the $V_{GS}$ values in Figure 7-3.

The P-channel FET must be biased so that its drain is negative with respect to its source and its gate must be positive with respect to the source. The drain current ($I_D$) will therefore flow in a direction that is opposite to the drain in an N-channel FET. However, the majority carriers (holes) within the P-type channel always move from the source to the drain just like the majority carriers (electrons) in the N-channel device.

# Self Test Review

1. Junction field effect transistors are usually made from germanium.

   a. True
   b. False

2. The construction of a junction FET usually starts with a section of lightly doped N-type or P-type material known as a _______________.

3. In a junction FET, current is made to flow through a ____________ that is implanted within the device.

4. The three leads associated with the junction FET are called the _____________, _____________, and _____________.

5. In most junction FET's, the _____________ and _____________ leads may be interchanged.

6. An N-channel junction FET utilizes an N-type channel and a P-type _______________.

7. In an N-channel FET the majority carriers are _______________.

8. In a P-channel FET the majority carriers are _______________.

9. The PN junction formed between the gate and source of a junction FET must always be _______________-biased.

10. The input resistance of a junction FET is much higher than the input resistance of a bipolar transistor.

    a. True
    b. False

11. The current flowing through the channel of an FET is referred to as _______________ current.

12. The current flowing through the channel of an FET can be effectively controlled by varying the reverse bias voltage applied to the gate and source leads.

    a. True
    b. False

13. The FET's gate-to-source voltage ($V_{GS}$) determines the size of the ______________ region that is formed within the FET's channel.

14 A junction FET operates in the ______________ mode.

15. A junction FET is biased so that an increase in $V_{GS}$ will result in a ______________ in $I_D$.

16. The value of $V_{GS\ (off)}$ for a given FET will be approximately equal to the FET's $V_p$ value.

    a. True
    b. False

17. The transconductance of an FET is determined by dividing a small change in $I_D$ by a small change in ____________.

18. A P-channel FET is properly biased if its gate is positive with respect to its source.

    a. True
    b. False

19. When the arrow in a junction FET symbol points inward, the device utilizes a channel that is made of ______________-type material.

20. A P-channel FET should be biased so that its drain is ______________ with respect to its source.

## Answers

1. b. False — Junction FET's are usually made from silicon.
2. substrate
3. channel
4. gate, source, and drain
5. source and drain
6. gate (or substrate)
7. electrons
8. holes
9. reverse
10. a. True
11. drain
12. a. True
13. depletion
14. depletion
15. decrease
16. a. True
17. $V_{GS}$
18. a. True
19. N
20. negative

# EXPERIMENT 8

## Junction FET Characteristics

*OBJECTIVE:* *To demonstrate the operation of a typical junction FET and determine its transconductance.*

### Introduction

In this experiment, you will examine the relationship between the gate-to-source voltage ($V_{GS}$) the drain current ($I_D$), and the drain-to-source voltage ($V_{DS}$) in an N-channel junction FET. You will first measure and record the corresponding $V_{GS}$, $I_D$, and $V_{DS}$ values; and then, plot these values to form a set of drain characteristics curves. Finally, you will complete the experiment by using these curves to determine the transconductance of the FET.

### Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — N-channel junction FET (417-291)
1 — 100 ohm resistor (brown, black, brown, gold)
2 — 1 kilohm resistor (brown, black, red, gold)

### Procedure

1. Turn on the Trainer and adjust the + voltage control for 15 volts DC. This voltage is applied to the 1 kilohm potentiometer (designated $R_5$). You will use this potentiometer to control the drain-to-source voltage ($V_{DS}$). The drain current flowing through the FET must pass through a 100 ohm resistor ($R_4$). As you proceed through this experiment, you will periodically measure the voltage across $R_4$, and then calculate $I_D$.

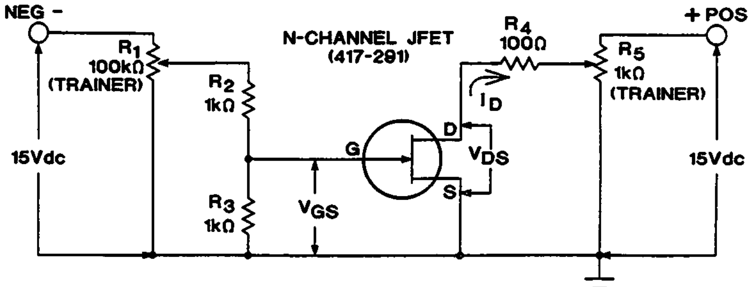


Figure 7-5
Experimental Circuit

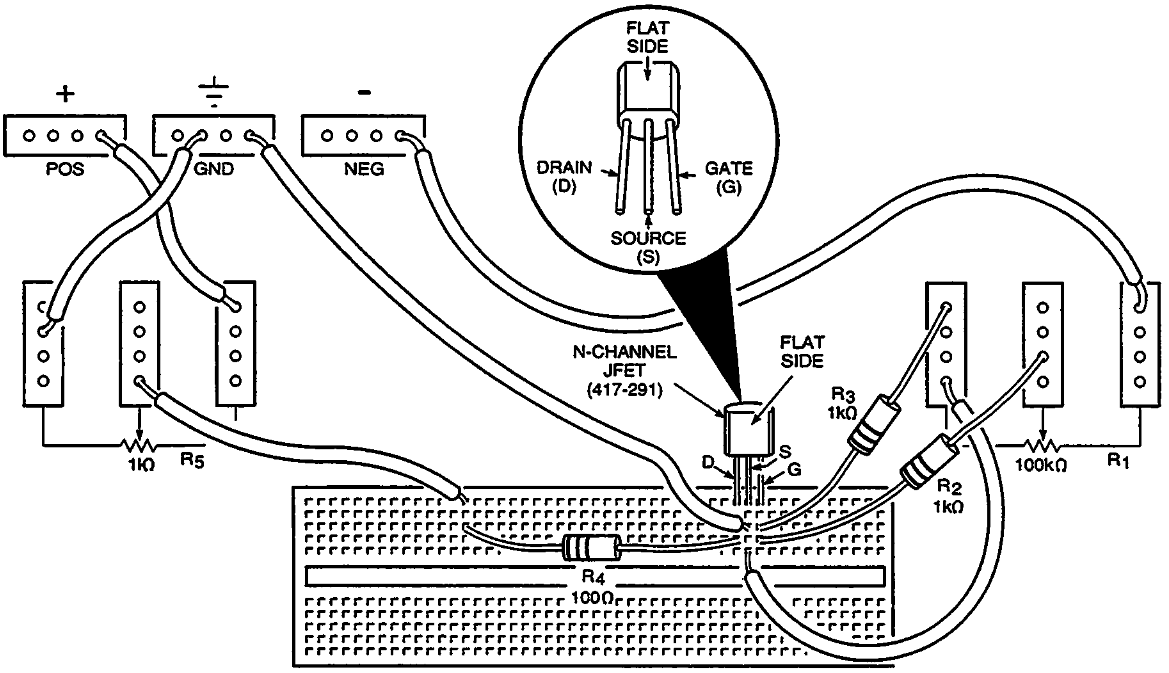


Figure 7-6
Wiring diagram for experimental circuit.

2. Adjust the – voltage control for –15 volts. This voltage is applied to the 100 kilohm potentiometer (designated $R_1$). You will use this potentiometer to control the voltage applied to the two 1 kilohm resistors ($R_2$ and $R_3$). These resistors will divide the voltage taken from $R_1$ in half. The reduced voltage which appears across $R_3$ serves as the gate-to-source bias voltage ($V_{GS}$).

3. Turn the Trainer off. Construct the circuit shown in Figure 7-5 using the wiring diagram shown in Figure 7-6. The leads of the N-channel junction FET (417-291) are identified in Figure 7-6.

4. Turn potentiometers $R_1$ and $R_5$ for minimum voltage and then turn on your Trainer.

5. Leave potentiometer $R_1$ so that the gate-to-source voltage ($V_{GS}$) applied to the FET will be zero. Adjust potentiometer $R_5$ until the drain-to-source voltage ($V_{DS}$) is equal to 1 volt.

6. Now measure the voltage across $R_4$. Use this voltage reading and the resistance of $R_4$ (100 ohms) to calculate the FET's drain current ($I_D$). Record your calculated $I_D$ value (in milliamperes) immediately below the $V_{DS}$ value of 1 volt in the table provided in Figure 7-7.

| $V_{GS} = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $V_{DS}$ (volts) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $I_D$ (milliamperes) | | | | | | | | |

Figure 7-7
Table for recording $I_D$ values when $V_{GS}$ equals zero.

7. Now complete the table in Figure 7-7 by adjusting $V_{DS}$ to the remaining values indicated and recording the corresponding values of $I_D$. To do this, simply repeat steps 5 and 6 for each value of $V_{DS}$ indicated. When you complete the table in Figure 7-7, you will have a permanent record of $I_D$ values over a range of $V_{DS}$ values when $V_{GS}$ is equal to zero.

8. Now adjust potentiometer $R_1$ so that $V_{GS}$ is equal to –0.25 volts.

9. Without disturbing the setting of $R_1$, adjust potentiometer $R_5$ until $V_{DS}$ is equal to 1 volt.

10. Now measure the voltage across $R_4$. Use this voltage reading and the resistance of $R_4$ (100 ohms) to calculate $I_D$. Record your calculated $I_D$ value (in milliamperes) immediately below the $V_{DS}$ value of 1 volt in the table provided in Figure 7-8.

| $V_{GS} = -0.25$ volts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $V_{DS}$ (volts) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $I_D$ (milliamperes) | | | | | | | | |

Figure 7-8
Table for recording $I_D$ values when $V_{GS}$ equals –0.25 volts.

11. Now complete the table in Figure 7-8 by adjusting $V_{DS}$ to the remaining values indicated and recording the corresponding values of $I_D$. To do this, simply repeat steps 9 and 10 for each value of $V_{DS}$ indicated. When you complete the table in Figure 7-8, you will have a record of $I_D$ values over a range of $V_{DS}$ values when $V_{GS}$ is equal to –0.25 volts. A negative sign is placed before the $V_{GS}$ value to indicate that the gate is negative with respect to the source.

12. Adjust potentiometer $R_1$ so that $V_{GS}$ is equal to –0.5 volts (measure with your voltmeter).

13. Without disturbing the setting of $R_1$, adjust potentiometer $R_5$ until $V_{DS}$ is equal to 1 volt.

14. Measure the voltage across $R_4$. Then use this voltage reading and the resistance of $R_4$ (100 ohms) to calculate $I_D$. Record your calculated $I_D$ value (in milliamperes) immediately below the $V_{DS}$ value of 1 volt in the table provided in Figure 7-9.

| $V_{GS}$ = -0.5 volts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $V_{DS}$ (volts) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $I_D$ (milliamperes) | | | | | | | | |

Figure 7-9

Table for recording $I_D$ values when $V_{GS}$ equals –0.5 volts.

15. Now complete the table in Figure 7-9 by repeating steps 13 and 14 for the remaining $V_{DS}$ values indicated. When you complete this table, you will have a permanent record of $I_D$ values over a range of $V_{DS}$ values when $V_{GS}$ is equal to –0.5 volts.

16. Adjust potentiometer $R_1$ so that $V_{GS}$ is equal to –0.75 volts.

17. Without disturbing the setting of $R_1$, adjust potentiometer $R_5$ until $V_{DS}$ is equal to 1 volt.

18. Measure the voltage across $R_4$. Then use this voltage reading and the resistance of $R_4$ (100 ohms) to calculate $I_D$. Record your calculated $I_D$ value (in milliamperes) immediately below the $V_{DS}$ value of 1 volt in the table provided in Figure 7-10.

| $V_{GS} = -0.75$ volts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $V_{DS}$ (volts) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $I_D$ (milliamperes) | | | | | | | | |

Figure 7-10
Table for recording $I_D$ values when $V_{GS}$ equals −0.75 volts.

19. Complete the table in Figure 7-10 by repeating steps 17 and 18 for the remaining $V_{DS}$ values indicated. When complete this table, you will have a record of $I_D$ values over a range of $V_{DS}$ values when $V_{GS}$ is equal to −0.75 volts.

20. Now use the corresponding values of $V_{DS}$ and $I_D$ that you recorded in Figure 7-7 to plot the drain characteristic curves on the graph shown in Figure 7-11. Connect the various points plotted to form a continuous curve. Label this curve $V_{GS} = 0$.

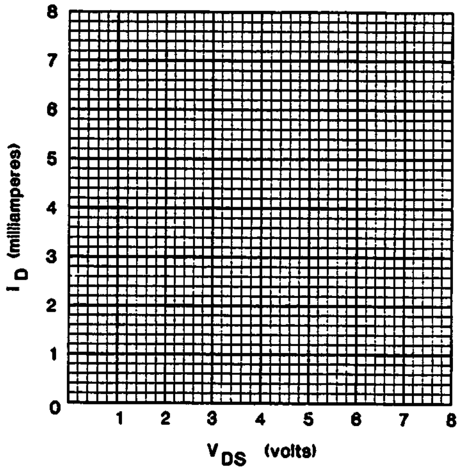


Figure 7-11
Chart for recording $V_{DS}$ and $I_D$ values.

21. Use the corresponding values of $V_{DS}$ and $I_D$ in Figure 7-8 to plot a second curve on the graph of Figure 7-11. Label this curve $V_{GS} = -0.25$ volts.

22. Use the corresponding values of $V_{DS}$ and $I_D$ in Figure 7-9 to plot a third curve on the graph of Figure 7-11. Label this curve $V_{GS} = -0.5$ volts.

23. Use the corresponding values of $V_{DS}$ and $I_D$ in Figure 7-10 to plot a fourth curve on the graph of Figure 7-11. Label this curve $V_{GS} = -0.75$ volts.

24. Now use the set of drain characteristic curves that you plotted in Figure 7-11 to determine the transconductance of the FET. Select a constant value of $V_{DS}$ that is well above the pinch-off voltage ($V_P$) of the device (possibly 6 or 7 volts) and observe the change in $I_D$ when $V_{GS}$ changes from 0 to −0.25 volts. This will insure that your transconductance value will be measured in the pinch-off region of the device. Use the transconductance equation given below and record your results in the space provided below.

$$\text{Transconductance} = \frac{\Delta I_D}{\Delta V_{GS}}$$

Transconductance = ——

25. Your experiment is now complete. Turn off your Trainer and test equipment.

## Discussion

In this experiment, you observed the relationship between $I_D$, $V_{DS}$, and $V_{GS}$ in an N-channel junction FET. You measured specific values of $I_D$ over a range of $V_{DS}$ values for several values of $V_{GS}$ and then plotted these values to produce a set of drain characteristic curves. The curves that you plotted should resemble the one shown in Figure 7-3 of this unit. However, your curves were not plotted for low values of $V_{DS}$ (between 0 and 1 volt) because of the difficulty in making accurate measurements within this region. Therefore, your curves do not begin at zero but instead begin at some value of $I_D$. Each curve should rise rapidly at first and then level off. At each higher value of $V_{GS}$, $I_D$ should level off at a correspondingly lower value.

Next you used your drain characteristic curves to determine the transconductance of the FET within its pinch-off region. You probably obtained a transconductance that was greater than 1500 micromhos (00015 mhos) but less than 5000 micromhos (0005 mhos). The exact transconductance that you obtained is not extremely important since this value usually varies widely even among FET's of the same type.

If you were not able to plot a drain characteristic curve for the highest value of $V_{GS}$ given ($V_{GS} = -0.75$ volts) because your $I_D$ values were all equal to zero, do not be concerned. This simply means that a $V_{GS}$ of −0.75 volts is high enough to cut off the drain current, or in other words, the $V_{GS(off)}$ rating for your FET was exceeded. This could occur if your FET has an unusually high transconductance. In such a case the FET would require only a very small $V_{GS}$ value to reduce its drain current to zero.

# THE INSULATED GATE FET

The gate and channel regions within a junction FET form a conventional PN junction and this junction is reversed-biased by connecting an external voltage between the FET's gate and source leads. This reverse bias voltage causes the FET to operate in the depletion mode and allows the device to have an extremely high input resistance. However, there is another type of FET which does not have a conventional PN junction that has to be reversed-biased. This device uses a metal gate which is electrically insulated from its semiconductor channel by a thin oxide layer and is therefore referred to as an *insulated gate* FET (IGFET). However, this device is also known as a *metal-oxide semiconductor* FET (MOSFET).

Unlike the junction FET which operates in only the depletion mode, the IGFET (MOSFET) is designed to operate in one of two distinct modes. The IGFET may be either a *depletion-mode* device or an *enhancement-mode* device.

You will now see how insulated-gate FET's are constructed and you will learn their important electrical characteristics.

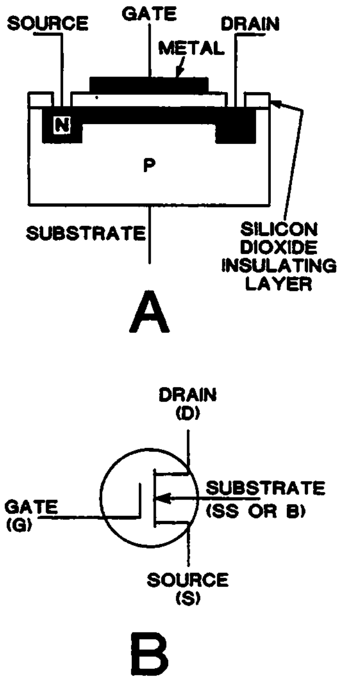


Figure 7-12
An N-channel depletion-mode IGFET (A) and its schematic symbol (B)

## Depletion-Mode Devices

A cross-sectional view of a depletion-mode IGFET (MOSFET) is shown in Figure 7-12A. This device is formed by implanting an N-type channel within a P-type substrate. A thin insulating layer (silicon dioxide) is then deposited on top of the device however the opposite ends of the N-type channel are left exposed so that wires or leads can be attached to the channel material. These two leads serve as the FET's source and drain. A thin metallic layer is then attached to the insulating layer so that it is directly over the N-type channel. This metal layer serves as the FET's gate and signals are applied to this metal gate through a suitable wire or lead. An additional lead is also connected to the substrate as shown.

It is important to note that the metal gate is insulated from the semiconductor channel by the layer of silicon dioxide. Therefore, the gate and the channel do not form a conventional PN junction. However, this insulated metal gate can be used to control the conductivity of the channel thus allowing the device to operate in a manner similar to that of a junction FET. In other words the insulated gate can deplete the N-type channel of majority carriers (electrons) when a suitable bias voltage is applied, even though a semiconductor junction does not exist between the gate and the channel.

The insulated gate FET shown in Figure 7-12A has an N-type channel; therefore its source and drain leads are biased in the same manner as the source and drain leads of an N-channel junction FET. In other words the drain is always made positive with respect to the source. The majority carriers (electrons) within the N-type channel therefore allow current to flow through the channel (from source to drain). This source-to-drain current (normally called drain current) is in turn controlled by a gate-to-source bias voltage just like a junction FET. When the gate-to-source voltage is equal to zero, a substantial drain current will flow through the device because a large number of majority carriers are present in the channel. If the gate is made negative with respect to the source, the channel becomes depleted of many of its majority carriers and drain current decreases just as it would in a junction FET. If this negative gate voltage is increased to a sufficiently high value, the drain current will drop to zero just like it does in the junction FET. There is one important difference between the operation of the N-channel insulated gate FET just described and the N-channel junction FET. The gate of the N-channel insulated gate FET can be made positive with respect to the FET's source. This must never occur in an N-channel junction FET because the PN junction formed by the gate and the channel would be forward biased.

The insulated gate FET can handle a positive gate voltage because the silicon dioxide insulating layer prevents any current from flowing through the gate lead. The FET's input resistance therefore remains at a high value. Also, when a positive voltage is applied to the gate, more majority carriers (electrons) are drawn into the channel thus enhancing the conductivity of the channel. This is exactly opposite to the action that takes place when the gate receives a negative voltage. A positive gate voltage can therefore be used to increase the FET's drain current while a negative gate voltage will decrease the drain current.

Since a negative gate voltage is required to deplete the N-channel insulated gate FET just described (and therefore reduce its drain current), this FET is called a depletion-mode device. Like the junction FET described earlier (also a depletion-mode device) this FET conducts a substantial amount of drain current when its gate voltage is equal to zero. Therefore, all depletion-mode devices (either junction or insulated gate types) are said to be normally-conducting or *normally-on* when their gate voltages are zero.

The N-channel, depletion-mode IGFET just described is often represented by the schematic symbol shown in Figure 7-12B. Notice that the gate (G) lead is separated from the source (S) and drain (D) leads. Also the arrow on the substrate (often designated as SS or B) lead points inward to represent an N-channel device. Some FET's are constructed so that the substrate is internally connected to the source lead and a separate substrate lead is not used. In such a case, the symbol can be drawn accordingly so that it shows the substrate connected to the source.

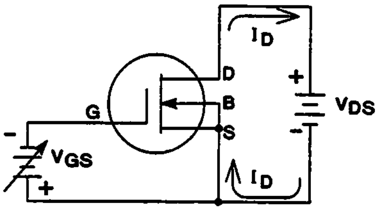


Figure 7-13
A properly biased N-channel depletion-mode IGFET

A properly biased N-channel, depletion-mode IGFET is shown in Figure 7-13. Notice that this FET is biased in the same manner as an N-channel junction FET. The drain-to-source voltage ($V_{DS}$) must always be applied so that the drain is positive with respect to the source as shown. However, the gate-to-source voltage ($V_{GS}$) can be applied as shown or the polarity of $V_{GS}$ can be reversed. In other words the gate can be made either negative or positive with respect to the source. Also notice that the substrate (B) has been externally connected to the source (S). The substrate is usually connected to the source (internally or externally) but this is not always the case. In some applications the substrate may be connected to the gate or to other points within the FET's respective circuit. These various circuit arrangements will be described later.

The relationship between $V_{GS}$, $V_{DS}$, and $I_D$ in a particular N-channel IGFET can be determined by examining the FET's drain characteristic curves. A typical set of curves are shown in Figure 7-14. Such curves could be plotted by using a circuit like the one shown in Figure 7-13. The curves are formed by adjusting $V_{GS}$ to various values and then observing the relationship between $I_D$ and $V_{DS}$ for each $V_{GS}$ value. If you compare these curves with the junction FET curves shown in Figure 7-3 you will find that they are similar. The basic difference is that positive

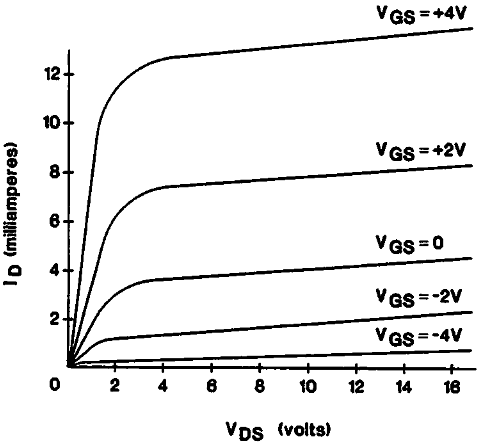


Figure 7-14
Typical drain characteristic curves for an N-channel depletion-mode IGFET

as well as negative $V_{GS}$ values are plotted in Figure 7-14 and only negative $V_{GS}$ values are plotted in Figure 7-3.

The curves for the depletion-mode IGFET have the same general shape as the curves for the junction FET. In other words each curve rises rapidly and then levels off at a specific pinch-off voltage ($V_P$). The depletion-mode IGFET (like the junction FET) is usually operated above pinch-off where $I_D$ is relatively constant with changes in $V_{DS}$. In this region the device may be used as a highly efficient voltage amplifier. However, there are some applications where the device is operated below pinch-off. Below pinch-off, the FET's drain current ($I_D$) varies over a wide range and at an almost linear rate as $V_{DS}$ changes. This means that the resistance of the device can be varied over a wide range by an input gate voltage thus making the device useful as a voltage controlled resistor. Furthermore, the N-channel depletion mode IGFET is usually operated with a slightly negative gate-to-source bias voltage ($V_{GS}$) but in some cases $V_{GS}$ may be equal to zero. This means that an input ac signal voltage can be applied to the FET's gate (in series with $V_{GS}$) so that the resulting $V_{GS}$ is effectively made to vary in value but remain negative or effectively vary between positive and negative values.

Depletion-mode IGFET's may also be constructed in a manner which is exactly opposite to the N-channel device shown in Figure 7-12. In other words they may utilize P-type channels which are implanted within N-type substrates. Such devices are generally referred to as P-channel, depletion-mode IGFET's. These P-channel devices operate in basically the same manner as the N-channel devices just described, however, the majority carriers within the P-type channel are holes instead of electrons. Furthermore, the drain lead of the P-channel device must be made negative with respect to its source lead so that its drain-to-source voltage ($V_{DS}$) is exactly opposite to the $V_{DS}$ applied to the N-channel device in Figure 7-13. This means that the drain current ($I_D$) must also flow in the opposite direction. However, the gate may be made negative or positive with respect to the source as with the N-channel device. The drain characteristic curves for the P-channel depletion-mode IGFET are therefore similar to the curves for an N-channel device but of course the polarities of the voltages and currents involved are reversed.

The schematic symbol for a P-channel, depletion-mode IGFET is shown in Figure 7-15. Notice that the only difference between this symbol and the symbol for an N-channel device is the direction of the arrow. In this symbol the arrow points outward to indicate that a P-type channel is used.

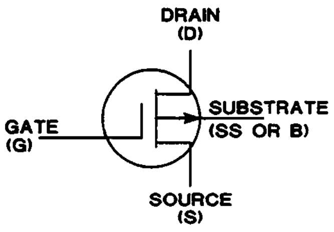


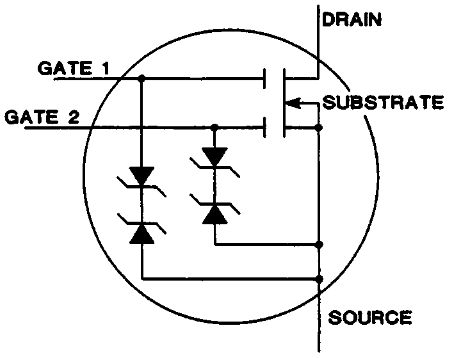
Figure 7-15
Schematic symbol for a
P-channel depletion-mode IGFET

Both the N-channel and P-channel, depletion-mode IGFET's may be constructed in a symmetrical manner (like the junction FET) so that their source and drain leads may be interchanged. However, these devices are sometimes constructed so that their gates are offset form their drain and improves the operation of the device in certain applications. When using this type of FET, the source and drain leads should not be interchanged.

One widely used variation of the depletion mode device is a dual gate unit. A dual gate IGFET (MOSFET) has two gate elements each of which can be used separately to control the drain current. Such devices are used as mixers or as gain controlled amplifiers in radio and TV receivers.

Figure 7-16 shows the symbol for an N-channel dual gate IGFET (MOSFET). Note the back-to-back zener diodes connected between each gate and the source. These are used to protect against electrostatic charges that can damage the device. The substrate is internally connected to the source. Each gate has equal control over the drain current.

Figure 7-16
A dual gate depletion mode IGFET with gate protection diodes



## Enhancement-Mode Devices

As explained previously, the depletion-mode IGFET is a normally-on device and therefore conducts a substantial drain current ($I_D$) when its gate-to-source voltage ($V_{GS}$) is zero. Although this type of FET is useful in many applications, there are also certain applications where a *normally-off* device is required. In other words it is often useful to have a device that conducts negligible $I_D$ when $V_{GS}$ is equal to zero but will allow $I_D$ to flow when a suitable $V_{GS}$ value is applied.

An IGFET which will function as a normally-off device can be constructed as shown in Figure 7-17A. This device is similar to the depletion-mode IGFET but it does not have a conducting channel which is imbedded in the substrate material. Instead, the device has source and drain regions which are diffused separately into the substrate. The device in Figure 7-17A has a P-type substrate and N-type source and drain regions; however, the exact opposite arrangement can also be used. The metal gate is separated from the substrate material by a silicon dioxide insulating layer as shown and the various leads are attached to the device so that it has basically the same lead arrangement as the depletion-mode device.

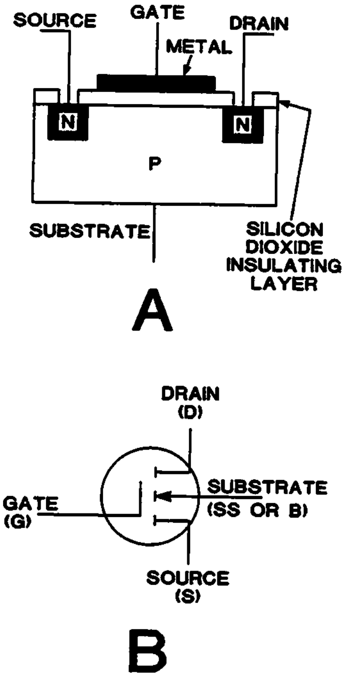


Figure 7-17
An N-channel enhancement-mode IGFET and its schematic symbol

The device shown in Figure 7-17A must be biased so that its drain is made positive with respect to its source. With only a drain-to-source voltage ($V_{DS}$) applied, the FET does not conduct a drain current ($I_D$). This is because no conducting channel exists between the source and drain regions. However, this situation can be changed with the application of a suitable gate voltage. When the gate is made positive with respect to the source, electrons are drawn towards the gate. These electrons accumulate under the gate where they create an N-type channel which allows current to flow from source to drain. When this positive gate voltage increases, the size of the channel also increases thus allowing even more drain current to flow.

The action just described is similar to the action that takes place in a charging capacitor. The metallic gate and the substrate act like the upper and lower plates of a capacitor and the insulating layer acts like the dielectric. The positive gate voltage simply causes the capacitor to charge and a negative charge builds up on the substrate side of the capacitor. The positive gate voltage effectively induces an N-type channel between the source and drain regions which sustains a drain current. Furthermore, an increase in gate voltage tends to *enhance* the drain current. For these reasons the device in Figure 7-17A is commonly referred to as an N-channel, enhancement-mode IGFET (MOSFET).

The gate of the N-channel, enhancement mode IGFET can also be made negative with respect to its source. However, a negative gate voltage will not affect the operation of the FET since it is a normally-off (non-conducting) device. In other words, the FET's drain current is normally equal to zero and therefore cannot be further reduced by application of a negative gate voltage.

The schematic symbol for an N-channel, enhancement-mode IGFET is shown in Figure 7-17B. Notice that this symbol is similar to the symbol for the N-channel depletion-mode IGFET shown in Figure 7-12B. The only difference is the use of an interrupted line instead of a solid line to interconnect the source, drain, and substrate regions. The solid line is used to identify the normally-on condition of the depletion-mode device while the interrupted line is used to identify the normally-off condition of the enhancement-mode device. In each symbol the arrow points inward to indicate that each device has an N-type channel.

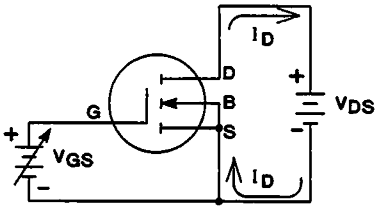


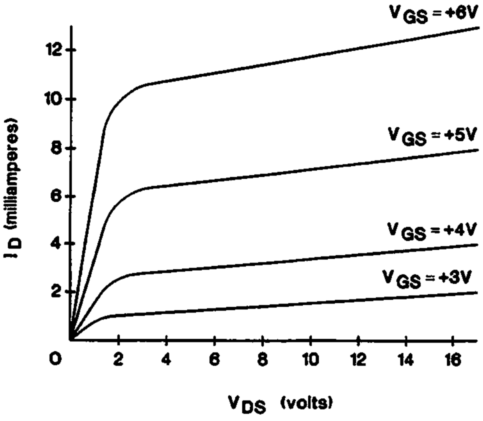
Figure 7-18
A properly biased N-channel enhancement-mode IGFET.

A properly biased N-channel, enhancement-mode IGFET is shown in Figure 7-18. Notice that the FET's drain is made positive with respect to its source by the drain-to-source voltage ($V_{DS}$). However, the gate is also made positive with respect to the source by the gate-to-source voltage ($V_{GS}$). Only when $V_{GS}$ increases from zero and applies a positive voltage to the gate, will a substantial amount of drain current ($I_D$) flow. The substrate is normally connected to the source as shown but in special applications the substrate and source may be at different potentials.

The relationship between $V_{GS}$, $V_{DS}$, and $I_D$ in a particular N-channel, enhancement-mode IGFET can be determined by examining the FET's drain characteristic curves. A typical set of these curves is shown in Figure 7-19. Such curves are usually plotted by biasing the FET as shown in Figure 7-18. Then $V_{GS}$ is adjusted to various values and the relationship between $I_D$ and $V_{DS}$ is observed for each $V_{GS}$ value. These curves are similar to the curves for an N-channel, depletion-mode IGFET but only positive values of $V_{GS}$ are plotted instead of both positive and negative values. Notice that for each higher positive value of $V_{GS}$, $I_D$ rises to a correspondingly higher value and then levels off.

Although it is not apparent in Figure 7-19, $V_{GS}$ must exceed a certain *threshold* voltage (usually one volt or more) before the N-type channel induced within the FET is great enough to support an appreciable $I\Delta$. Any $V_{GS}$ value below this threshold cannot cause the FET to conduct and the device will effectively act as if its gate voltage was equal to zero. The N-channel enhancement-mode IGFET therefore normally operates with a positive gate bias that is greater than its threshold voltage. With proper biasing the enhancement-mode device makes an excellent switch. The device can be turned on by a sufficiently high gate voltage and turned off when the gate voltage drops below the threshold level. The inherent threshold of the device therefore provides a highly desirable region of noise immunity which prevents low or intermediate input voltages (be-

Figure 7-19
Typical drain characteristic curves for an N-channel enhancement-mode IGFET



low threshold) from falsely triggering the device on. This characteristic makes the enhancement mode device perfect for digital applications involving logic and switching functions. The enhancement mode IGFET is the basic component used in many large scale digital integrated circuits.

An enhancement-mode IGFET may also be constructed in a manner that is exactly opposite to the device shown in Figure 7-17A. In other words the device may have P-type source and drain regions which are implanted in an N-type substrate. This type of FET must be operated with a negative gate voltage so that holes (instead of electrons) are attracted toward the gate to form a P-type channel. Such a device is referred to as a P-channel, enhancement-mode IGFET (MOSFET). The P-channel device functions in basically the same manner as the N-channel device even though holes instead of electrons are used to support drain current through the device. The P-channel device requires bias voltages (V GS and $V_{DS}$) that are opposite to those shown in Figure 7-18. Also, the drain characteristic curves for the P-channel device have the same general shape as the curves for the N-channel device although the polarities of the voltages and currents involved are reversed.

The symbol for a P-channel, enhancement-mode IGFET is shown in Figure 7-20. This symbol closely resembles the symbol for the N-channel device. The only difference is the direction of the arrow which in this case points outward to identify the P-type channel that is induced within the device. The enhancement-mode device is usually constructed in a symmetrical manner just like the junction FET. This means that the source and drain leads can usually be reversed or interchanged.

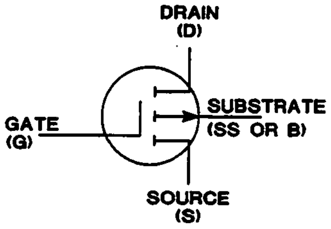


Figure 7-20
Schematic symbol for a P-channel, enhancement-mode IGFET

# CMOS

CMOS (Complementary-Metal-Oxide-Semiconductor) components were developed using a logical expansion of the MOS technology that followed the development of TTL logic (bipolar components). As you know, bipolar devices (PNP and NPN transistors) were combined to make a circuit called a complementary pair. MOS devices were also developed as P-type and N-type components, and then externally combined to form a complementary pair, or a complementary amplifier. CMOS technology combines the N-type and P-type FETs together into a single unit that is a complementary pair. This is shown in Figure 7-21.

Compared to the bipolar technology, CMOS technology requires less space, reduces the power dissipated, and the noise generated internally.

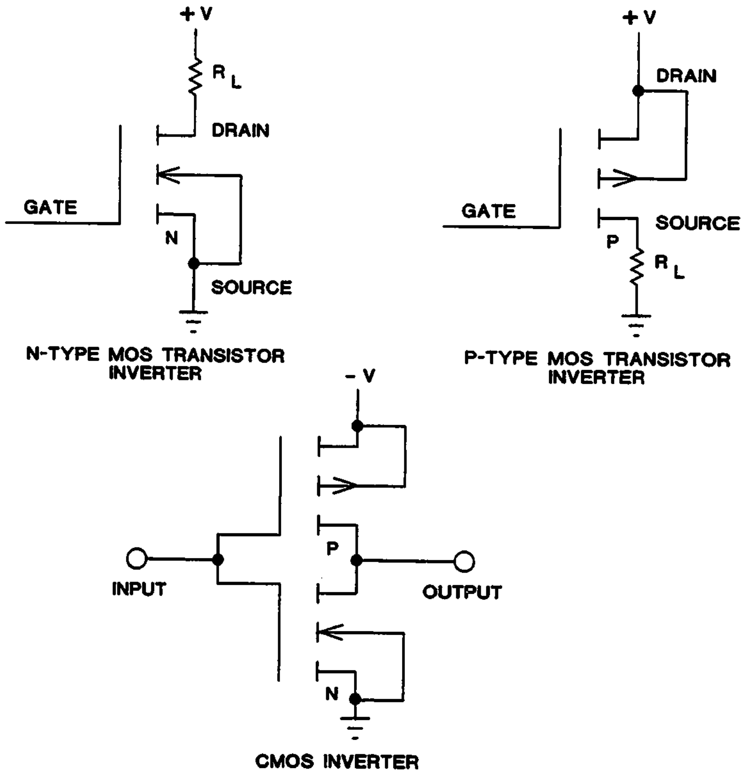


Figure 7-21
Schematic diagrams of inverter circuits.

Remember, that the FET operates at slower speeds than a bipolar device. The FET is also limited to lower power applications than the bipolar components. Because of their similar technologies, any of the functions that can be performed by conventional FETs can be accomplished with CMOS devices. However, CMOS operates at slower speeds than the P-type or N-type FETs. This is because the propagation delay (the time it takes a signal to travel through a device) is longer in a CMOS device.

In addition to slow operating speeds (approximately 5 million changes per second) and low power capabilities, CMOS components are more susceptible to damage from static electricity than either TTL or FET components. Static electricity can build up on your fingers to a high enough potential to destroy a CMOS component. Therefore, special handling and packaging are required.

# Safety Precautions

When using any MOSFET (IGFET, CMOS, etc.) device, certain precautions must be observed. As with any solid-state component, it is necessary to check the manufacturer's maximum rating so that the device is not damaged by excessive high operating voltages or currents. However, it is particularly important to observe the FETs maximum allowable gate-to-source voltage ($V_{GS}$). A MOSFET can accept only a limited range of $V_{GS}$ values because of the extremely thin silicon dioxide insulating layer that separates its gate and channel. If $V_{GS}$ is increased too much, the thin insulating layer will be punctured, and the device ruined. In fact, the insulating layer is so sensitive that it can even be damaged by static charges that build up on the FETs leads. For example, the electrostatic charges on your fingers can be transferred to the FETs leads while handling the device or when mounting it in a circuit. The device can therefore be ruined before it is used. To avoid this type of damage, manufactures usually ship these devices with their leads shorted together so that static charges cannot build up between the leads. The leads may be wrapped with a shorting wire, or inserted within a shorting ring, or pressed into a conducting foam material, or simply taped together.

These shorting devices should not be removed until the FET is completely installed into its repective circuit. Components should not be installed or removed when power is applied to the circuit.

Many MOSFETs are protected by zener diodes, which are electrically connected between each insulated gate and the source inside the device. These diodes offer protection against static discharge and in-circuit transients without the need for external shorting mechanisms. Devices which do not include gate-protection diodes can be handled safely if the following basic precautions are taken:

# Safety Notes

1. Prior to connection in a circuit, all leads of the device should be kept shorted together either by the use of a metal shorting ring attached to the device by the manufacturers, or by their insertion into some kind of conductive material. Aluminum foil is sometimes used, since it can be readily torn away after the device is installed.

2. When devices are removed by hand from their carriers, the hand being used should be grounded by a suitable means, for example, with a metallic wristband.

3. Tips of soldering irons should be grounded before soldering the device.

4. The device should never be inserted into, or removed from circuits with power on.

## Self Test Review

21. An insulated gate FET has a metallic gate which is electrically insulated from its semiconductor ______________.

22. An IGFET may be designed to operate in the depletion mode or the enhancement mode.

    a. True
    b. False

23. A depletion-mode IGFET can be either a ______________-channel or ______________-channel device.

24. A depletion-mode IGFET operates similar to a junction FET but does not have a PN ______________.

25. The majority carriers within an N-channel, depletion-mode IGFET are ______________.

26. The N-channel, depletion-mode IGFET is always biased so that its drain is ______________ with respect to its source.

27. When the N-channel, depletion-mode IGFET's gate-to-source voltage is equal to zero, drain current will flow through the device.

    a. True
    b. False

28. When the gate of an N-channel, depletion-mode IGFET is made increasingly negative with respect to its source, the FET's drain current will ______________.

29. If the gate of an N-channel, depletion-mode IGFET is made increasingly positive with respect to its source, the FET's drain current will ______________.

30. The relationship between $V_{GS}$, $V_{DS}$, and $I_D$ in a depletion-mode IGFET can be determined by examining the FET's ___________ ___________ curves.

31. A P-channel, depletion-mode IGFET requires bias voltages that are exactly ______________ to those used with an N-channel device.

32. The majority carriers within a P-channel, depletion-mode IGFET are ______________.

33. Depletion-mode IGFET's are often referred to as normally-____________ devices.

34. An enhancement-mode IGFET conducts only a negligible drain current when $V_{GS}$ is equal to ____________.

35. The enhancement-mode IGFET is said to be a normally-____________ device.

36. In an N-channel, enhancement-mode IGFET, a positive gate voltage will attract ____________ toward the gate to form an N-type channel.

37. In order to conduct drain current an N-channel enhancement-mode IGFET must have a ____________ gate voltage.

38. The metallic gate and the substrate in an enhancement-mode IGFET act like the plates of a ____________.

39. The drain current in an N-channel, enhancement-mode IGFET can be reduced below its normal value by application of a negative gate voltage.

    a. True
    b. False

40. The N-channel, enhancement-mode IGFET will normally be biased so that its drain is ____________ with respect to its source.

41. The drain characteristic curves for the N-channel, enhancement-mode IGFET resemble the curves for an N-channel, depletion mode IGFET, however the enhancement-mode curves are only plotted for ____________ values of $V_{GS}$.

42. In order for an enhancement mode IGFET to conduct a substantial drain current, $V_{GS}$ must exceed a certain ____________ voltage.

43. In a P-channel, enhancement-mode IGFET, a negative gate voltage will attract ____________ toward the gate to form a P-type channel.

44. When using any type of insulated gate FET it is necessary to insure that static charges do not build up on the component's leads so that the highly sensitive ____________ layer will not be damaged.

45. Some IGFETs are protected from electrostatic charges by internally connected ____________ ____________.

## Answers

21. channel
22. a. True
23. P-channel or N-channel
24. junction
25. electrons
26. positive
27. a. True
28. decrease
29. increase
30. drain characteristic
31. opposite
32. holes
33. on (or conducting)
34. zero
35. off
36. electrons
37. positive
38. capacitor
39. b. False — the N-channel device is normally off, therefore $I_D$ cannot be further reduced by a negative gate voltage.
40. positive
41. positive
42. threshold
43. holes
44. insulating
45. zener diodes

# FET CIRCUIT ARRANGEMENTS

Like bipolar transistors, FET's are used primarily to obtain amplification; and like bipolar transistors, FET's can be connected in three different circuit arrangements. These three configurations are commonly referred to as *common-source, common-gate,* and *common-drain* circuits. These configurations are valid for both JFETs and IGFETs.

## Common-Source Circuits

The common-source circuit is the most widely used FET circuit arrangement. It is comparable to the common-emitter (bipolar-transistor) circuit that was described in a previous unit. A basic common-source configuration is shown in Figure 7-22. Notice that the input signal is applied between the gate and source leads of the FET, and the output signal appears between the drain and source leads. The source is therefore common to both the input and the output.

An N-channel junction FET is used in Figure 7-22. Therefore, it must be biased so that its gate is negative with respect to its source. The gate-to-source bias voltage is provided by an external voltage source (designated as $V_{GG}$) which is in series with resistor $R_G$. Therefore, $V_{GG}$ is not applied directly to the gate and source leads. However, the FET's gate-to-channel PN junction is reverse-biased so that only an insignificant leakage current can flow through the gate lead and $R_G$. The voltage across $R_G$ is therefore almost zero and the full value of $V_{GG}$ is effectively placed across the gate and source leads. Normally, $R_G$ has a high value of resistance (often more than 1 megohm) so that the resistance at the input of the circuit will remain high. A low value of resistance could reduce the input resistance of the circuit since $R_G$ is effectively in parallel with the FET's high input (gate-to-source) resistance.

The external bias voltage ($V_{GG}$) is adjusted so that a specific value of drain current ($I_D$) will flow through the FET and the device will operate within its pinch-off region. The value of $I_D$ is also controlled (but to a lesser extent) by external voltage source $V_{DD}$. This voltage source supplies the necessary drain-to-source operating voltage for the FET but another resistor ($R_L$) is inserted between $V_{DD}$ and the FET's drain lead. The drain current flowing through the FET therefore flows through $R_L$ thus causing a voltage to appear across this resistor. A portion of the voltage ($V_{DD}$) is therefore dropped across the FET and the rest appears across $R_L$.

Since the FET is controlled by an input voltage (not an input current), the common-source arrangement is used to obtain voltage amplification. For example, an ac input voltage will alternately aid and oppose the input bias voltage ($V_{GG}$) so that the FET's gate-to-source voltage will vary with the changes in input voltage. This will, in turn, cause the FET to alternately conduct more and less drain current. The FET, therefore, acts like a variable resistor in series with a fixed resistor ($R_L$) and the resistance of the FET is effectively controlled by the input signal voltage. As the FET conducts more and less drain current, its drain-to-source voltage varies accordingly to produce an output voltage that changes in response to the input voltage. However, making $R_L$ relatively large (often more than 10 kilohms) and be biasing the FET so that its drain-to-source resistance is also high, the changes in drain current (even when small) can produce an output signal voltage that is much higher than the input signal voltage.

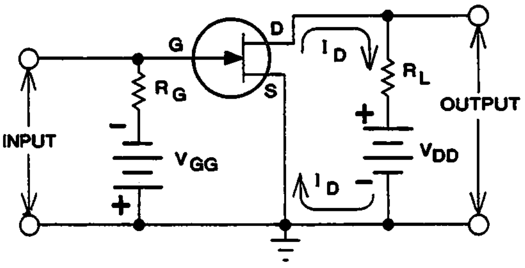


Figure 7-22
Basic common-source circuit

In addition to providing voltage amplification, the common-source circuit has another desirable feature. Due to the extremely high gate-to-source resistance of the FET, the circuit has a very high input resistance even though input resistor $R_G$ is used. This means that the common-source circuit will usually have a minimum loading effect on its input signal source. Common-source circuits are therefore widely used in digital or computer applications where a number of circuit inputs must often be connected to the output of one circuit without affecting its operation.

The output resistance of the common-source circuit is lower than its input resistance because of the lower source-to-drain resistance of the FET. However, the circuit still has a moderately high output resistance since the FET is usually biased to conduct a relatively low drain current which is often not more than a few milliamperes.

The common-source circuit may also be used to amplify both low frequency and high frequency ac signal voltages as well as a wide range of dc voltages. For example, this circuit is often used in the input rf amplifier stage of radio receivers to amplify the wide range of high frequency input signals which can vary widely in amplitude. The circuit is also used in electronic test instruments such as solid-state voltmeters or multimeters. In these applications, it must amplify a wide range of dc and ac voltages and at the same time present a high input resistance to prevent undesirable loading of the circuits being tested.

Although the common-source circuit shown in Figure 7-22 is formed with an N-channel junction FET, the same basic circuit can be formed with a P-channel JFET if the polarities of the bias voltages are reversed. Also, this basic circuit can be formed with depletion-mode and enhancement-mode IGFET's. When these insulated-gate devices are used, the additional substrate leads are generally connected to their respecting source leads or to circuit ground. Also, it is important to note that depletion-mode IGFET's can accept positive and negative gate voltages and may be operated with zero gate bias, while enhancement-mode devices require a gate bias voltage in order to conduct. This means that slightly different biasing arrangements must be used to accommodate the various FET types when the basic common-source arrangement is used.

## Common-Gate Circuits

The common-gate circuit may be compared to the common-base bipolar transistor circuit, because the electrode which has primary control over the FET's conduction (its gate) is common to the circuit's input and output. A basic common-gate circuit is shown in Figure 7-23. An N-channel JFET is used in this circuit. Therefore it requires the same basic operating voltages as the N-channel device in the common-source circuit previously described. However, in this circuit the input gate-to-source bias voltage is provided by voltage $V_{SS}$ and resistor $R_S$, while the output portion of the circuit is biased by voltage $V_{DD}$ and resistor $R_L$.

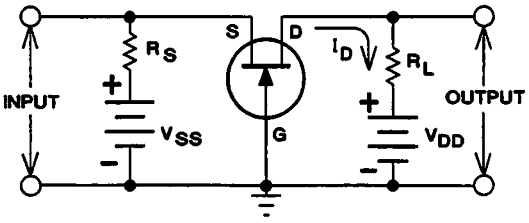


Figure 7-23

Basic common-gate circuit

An ac input voltage will effectively vary the FET's gate-to-source voltage and cause variations in its conduction. The FET's drain current flows through $R_L$ and variations in this current produce voltage changes across the FET which follow the input signal. The voltage developed across the drain and gate leads of the FET serves as the output.

Like the common-source circuit, the common-gate circuit provides voltage amplification; however, the voltage gain of the common-gate arrangement is lower. The input resistance of the circuit is also low since current flows through the input source lead; however, the output resistance of the circuit is relatively high. This makes the common-gate circuit suitable in applications where a low resistance source must supply power to a high resistance load. When inserted between the source and load, the common-gate circuit effectively matches the low and high resistances to insure an efficient transfer of power.

Although it has a low voltage gain, the common-gate circuit is often used to amplify high frequency ac signals. This is because of the circuit's low input resistance and because the circuit inherently prevents any portion of its output signal from feeding back and interfering with its input signal. The circuit is therefore inherently stable at high frequencies and additional components are not required to prevent interference between input and output signals. Such stabilizing components are often required with the common-source circuit to insure reliable operation.

The common-gate circuit may also be formed with a P-channel JFET or with depletion-mode or enhancement-mode IGFETs. However, when an IGFET is used, the substrate lead is usually connected directly to the gate or to circuit ground, and additional components may be required to properly bias the device.

# Common-Drain Circuits

The common-drain circuit is similar to the common-collector bipolar transistor circuit. This basic FET circuit arrangement is shown in Figure 7-24. The N-channel JFET in this circuit receives its gate-to-source bias voltage from $V_{GG}$ and $R_G$. The output (drain-to-source) portion of the FET is biased by $V_{DD}$ and $R_L$. The input signal voltage is effectively applied between the FET's gate and drain even though it appears to be applied to the gate and source leads. The output signal is effectively developed across the FET's source and drain even though it appears to be taken directly from $R_L$. This is because voltage source $V_{DD}$ is effectively a short as far as the input and output signals are concerned. Therefore, $V_{DD}$ effectively grounds the FET's drain and makes it common to both the input and the output signals.

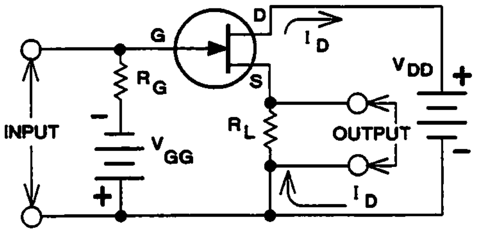


Figure 7-24
Basic common-drain or source follower circuit

The common-drain circuit cannot provide voltage amplification. The voltage appearing at the source of the FET tends to track, or follow, the voltage at the gate, and for this reason, the circuit is often referred to as a *source-follower*. However the output source voltage is always slightly less than the input gate voltage thus causing the circuit to have a voltage gain that is less than one, or unity.

The input resistance of the common-drain circuit is extremely high. In fact, this circuit has a higher input resistance than a common-source or a common-gate configuration. However, the output resistance of this circuit is low. This makes the common-drain circuit suitable for coupling a high resistance source to a low resistance load so that an efficient transfer of power can take place.

The common-drain configuration can also be formed with the P-channel JFETs as well as depletion-mode and enhancement-mode IGFETs.

## Self-Test Review

46. The most widely used FET circuit arrangement is the ______________-______________ circuit.

47. Only N-channel JFET's can be used to form the three basic FET circuit arrangements.

    a. True
    b. False

48. The FET circuit arrangement that has the highest voltage gain is the ______________-______________ circuit.

49. The FET circuit arrangement that inherently provides the most stable operation at high signal frequencies is the ______________-______________ circuit.

50. The common-drain arrangement provides a voltage gain of less than ______________.

51. The common-drain arrangement is sometimes referred to as a ______________-______________.

### Answers

46. common-source
47. b. False
48. common-source
49. common-gate
50. one or unity
51. source-follower

# UNIT SUMMARY

The field effect transistor can provide amplification of electronic signals like a conventional bipolar transistor. In fact, the FET is even more efficient than a bipolar transistor in certain applications. However, the FET operates on an entirely different principle and it has certain unique features that a bipolar transistor does not possess.

The operation of the bipolar transistor depends on the movement of both majority and minority carriers, but the FET operates only with majority carriers. The majority carriers flow through a semiconductor channel which is implanted within an oppositely doped substrate. The opposite ends of the channel are referred to as the FET's source and drain and may be compared to the emitter and collector in a bipolar transistor. The substrate itself is used to control the movement of majority carriers through the channel and it is referred to as the FET's gate. This gate may be compared to the base of a bipolar transistor. The device just described is more appropriately referred to as a junction FET since the substrate (gate) and the semiconductor channel form a PN junction.

Junction FET's may be constructed with an N-type channel and a P-type substrate or with a P-type channel and an N-type substrate. These devices are generally referred to as just N-channel or P-channel junction FET's respectively. The substrate (gate) to channel junction of the FET must always be reverse-biased so that a depletion region will form within the channel. By adjusting this reverse voltage, the size of the depletion region can be controlled. Since the size of the depletion region in turn controls the movement of the majority carriers through the channel, the gate-to-channel voltage can be used to effectively control the conduction of the FET.

The gate-to-channel junction is usually reverse-biased by applying the bias voltage between the FET's gate and source leads. In addition to this gate-to-source bias voltage, the FET also requires a source-to-drain bias voltage to force the majority carriers to move through the channel. The majority carriers will be electrons in an N-channel device and holes in a P-channel device but in either case only electrons leave the channel to flow through the external bias voltage source and return to the channel to complete the circuit. This external flow of electrons is referred to as the FET's drain current.

The FET's gate-to-source voltage therefore controls the depletion region within its channel and the size of this region in turn controls the FET's drain current. The junction FET is therefore said to operate in the depletion-mode.

The junction FET represents one basic type of FET. A second type, known as an insulated gate FET (IGFET) is also used. The insulated gate FET, also called a MOSFET, is constructed in a manner similar to the junction device but it utilizes a metallic gate which is electrically insulated from its semiconductor channel. The substrate is used only to support the entire structure. The substrate may be internally connected to the source or a separate electrical connection to the substrate may be provided. Insulated gate FET's are also available as N-channel and P-channel devices. Furthermore, they can be designed to operate in either the depletion mode like the junction FET or they can operate in the enhancement mode. The depletion-mode device operates in manners similar to the junction FET but its gate can be made either negative or positive with respect to its source. This is true if the device has either an N-type or a P-type channel. Furthermore, the depletion-mode device conducts drain current with a zero gate-to-source voltage just like the junction FET and is therefore referred to as a normally-on device.

The enhancement-mode device does not conduct drain current when its gate-to-source voltage is zero. Therefore, this type of insulated gate FET is often referred to as a normally-off device. A gate-to-source voltage must be applied to enhance conduction within the device. The gate of the enhancement-mode FET can be made either positive or negative with respect to its source. However, only one polarity of input gate voltage can be used to turn on the device and control its conduction. In general N-channel devices require positive gate voltages and P-channel devices require negative gate voltages.

Either type of FET can be connected in a common-source configuration so that the input signal is applied between its gate and source and the output signal voltage is developed between its drain and source. In this configuration the FET's reversed biased gate-to-channel junction gives the FET and the circuit a high input resistance. This arrangement also provides a substantial voltage gain. However the FET can be connected in a common-gate arrangement which has a somewhat lower voltage gain but provides more stable operation at high frequencies. A third basic configuration is also possible which is known as a common-drain or source-follower circuit. This arrangement provides no voltage gain but offers a very high input resistance and a low output resistance and can be used to insure maximum transfer of power from a high resistance source to a low resistance load.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. The junction FET has three leads which are referred to as the

   A. emitter, base, and collector.
   B. emitter, gate, and collector.
   C. source, gate, and collector.
   D. source, gate, and drain.

2. The junction FET operates in the

   A. enhancement mode
   B. depletion mode.
   C. combined depletion-enhancement mode.
   D. depleton or enhancement mode.

3. When a junction FET's gate-to-source voltage is zero and the FET's drain-to-source voltage is greater than the FET's pinch-off voltage, the device will

   A. conduct a substantial gate current.
   B. operate in the enhancement mode.
   C. conduct a substantial drain current.
   D. not conduct drain current.

4. The drain current produced by a properly biased N-channel junction FET will decrease when the FET's gate is

   A. made more negative with respect to its source.
   B. made less negative with respect to its source.
   C. shorted to its source to reduce $V_{GS}$ to zero.
   D. subjected to a potential which decreases the size of the depletion region.

5. The input resistance of a junction FET will remain extremely high as long as the FET's gate-to-channel junction is:

   A. biased so that the FET is operating in the enhancement mode.
   B. biased so that the gate current flows.
   C. forward-biased.
   D. reverse-biased.

6. The amplifying ability of an FET is expressed mathematically by the ratio

   A. $\Delta I_D/V_{DS}$ with $V_{GS}$ constant.
   B. $\Delta I_D/\Delta V_{GS}$ with $V_{DS}$ constant.
   C. $\Delta I_G/\Delta V_{GS}$ with $V_{DS}$ constant.
   D. $\Delta V_{GS}/\Delta I_D$ with $V_{DS}$ constant.

7. An insulated gate FET can be designed to operate in

   A. the enhancement mode only.
   B. the depletion mode only.
   C. either the depletion or the enhancement mode.
   D. both the depletion and enhancement modes simultaneously.

8. Reversing the polarity of the gate-to-source voltage applied to a properly biased depletion-mode insulated gate FET will

   A. lower the FET's input resistance.
   B. produce a change in the FET's drain current.
   C. permanently damage the FET.
   D. cause the FET to conduct drain current in the opposite direction.

9. The enhancement-mode insulated gate FET always conducts zero drain current when its

   A. gate-to-source voltage is below a certain threshold value.
   B. drain-to-source voltage is very high.
   C. gate current is equal to zero.
   D. gate-to-source voltage is greater than a certain threshold value.

10. The relationship between an FET's gate-to-source voltage, drain current, and drain-to-source voltage can be determined by examining the FET's

    A. source characteristic curves.
    B. gate characteristic curves.
    C. collector characteristic curves.
    D. drain characteristic curves.

11. FET's are often constructed in a symmetrical manner which makes it possible to interchange their

    A. source and gate leads.
    B. source and drain leads.
    C. gate and drain leads.
    D. gate and substrate leads.

12. The highest voltage gain is obtained when using an FET that is connected in a

    A. common-collector circuit.
    B. common-drain circuit.
    C. common-gate circuit.
    D. common-source circuit.

13. The current that flows in a junction FET when $V_{GS} = 0$ is commonly identified as

    A. $I_{GS}$
    B. $I_{DS}$
    C. $I_{DSS}$
    D. $I_S$

14. An IGFET is also referred to as a

    A. MOSFET
    B. NPN
    C. JFET
    D. PNP

15. If $V_{DS}$ is greater than $V_p$, the drain current in a JFET will

    A. be zero
    B. vary little with changes in $V_{DS}$
    C. be inversely proportional to gate voltage
    D. be proportional to gate voltage

# EXAMINATION ANSWERS

1. D — These three leads may be compared to the bipolar transistor's emitter, base and collector.

2. B — The junction FET's gate-to-channel junction is reversed-biased to create a depletion region within the FET's channel and this region effectively controls the FET's conduction.

3. C — Under these conditions the FET would be operating within its pinch-off region and its drain current would be relatively high.

4. A — The more negative gate voltage causes the depletion region to increase in size and reduce drain current.

5. D — The FET's gate-to-source bias voltage must cause the PN junction formed by the gate and the channel to be reversed-biased so that the gate current will be effectively zero and the input resistance will be high.

6. B — This ratio represents the FET's transconductance (gm).

7. C — Both depletion-mode and enhancement-mode insulated gate FET's are widely used.

8. B — Both positive and negative gate voltages can be applied to the depletion-mode IGFET to control its drain current without damaging the device or lowering its input resistance.

9. A — The gate-to-source voltage must exceed a certain threshold before the channel created within the enhancement-mode device is great enough to support an appreciable drain current.

10. D — The FET's drain characteristic curves serve the same basic purpose as the bipolar transistor's collector characteristic curves.

11. B — Although most FET's are symmetrical in construction, all of them are not and it is always best to check the manufacturer's specifications to be sure.

12. D — The common-source arrangement is the most popular FET configuration since it has the highest voltage gain and still offers a very high input resistance.

13. C — This is the drain current at $V_{GS} = 0$.

14. A — The term MOSFET is also used to identify an insulated gate field effect transistor.

15. B — The FET is operating in the pinch-off region when $V_{DS}$ is greater than $V_p$. Therefore $I_D$ remains almost constant as $V_{DS}$ is varied.

# Unit 8

# THYRISTORS

# CONTENTS

Introduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-3

Unit Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-4

Unit Activity Guide. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-5

Silicon Controlled Rectifiers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-6

  Basic Construction and Operation. . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-6

  V-I Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-8

  Practical Applications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-11

Experiment 9. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-17

Bidirectional Triode Thyristors. . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-24

  Basic Construction and Operation. . . . . . . . . . . . . . . . . . . . . . . . .Page 8-24

  V-I Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-26

  Applications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-27

Bidirectional Trigger Diodes. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-31

  Basic Construction and Operation. . . . . . . . . . . . . . . . . . . . . . . . .Page 8-31

  V-I Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-32

  Applications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-33

Unijunction Transistors. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-36

  Basic Construction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-36

  Operation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-38

  V-I Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-41

  Applications of the UJT . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-42

  Programmable UJT's. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-44

Experiment 10. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-50

Unit Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-56

Unit Examination . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-57

Examination Answers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .Page 8-59

# INTRODUCTION

The term *thyristor* defines a broad range of solid-state components which are used as electronically controlled switches. Each of these devices can switch between a conducting (on) state and a nonconducting (off) state to effectively pass or block electrical current. Futhermore, some thyristors are capable of switching currents flowing in one direction while others can switch currents flowing in either direction.

Thyristors are widely used in applications where dc and ac power must be controlled. These devices are often used to apply a specific amount of power to a load or to completely remove it from the load. However, they are also used to regulate or adjust the amount of power applied to a specific load. For example, a thyristor could be used to simply turn an electric motor on or off or it could be used to adjust the speed or torque of the motor over a wide operating range.

Thyristors should not be confused with bipolar transistors or field effect transistors (FET's). Although it is true that transistors and FET's can be used as electronic switches, these devices are not as efficient and they do not have the power handling capability of thyristors. Thyristors are devices that are used expressly for the purpose of controlling electrical power while transistors and FET's are primarily used to provide amplification.

A variety of thyristors are now available but some of these devices have similar or related characteristics. Most applications which involve power control are therefore handled with a few basic components. The thyristors that are most widely used are the *silicon controlled rectifier*, the *bidirectional triode thyristor*, the *bidirectional trigger diode*, the *unijunction transistor*, and the *programmed unijunction transistor*. You will examine each of these devices as you proceed through this unit. You will also demonstrate the operation of the silicon controlled rectifier and the unijunction transistor.

# UNIT OBJECTIVES

When you have completed this unit on thyristors you should be able to:

1. Describe the conditions necessary to turn on or turn off a silicon controlled rectifier.
2. Explain the difference between a silicon controlled rectifier's forward breakover and reverse breakdown voltage.
3. Name two applications of the silicon controlled rectifier.
4. Describe the conditions necessary to turn on or turn off a bidirectional triode thyristor.
5. Name two applications of the bidirectional triode thyristor.
6. Describe the basic operation of the bidirectional trigger diode.
7. Name the most important application of the bidirectional trigger diode.
8. Describe the conditions required to turn on a unijunction transistor.
9. Explain how a unijunction transistor exhibits a negative resistance once it is turned on.
10. Name two applications of the unijunction transistor.
11. Describe the difference between an ordinary unijunction transistor and a programmed unijunction transistor.

# UNIT ACTIVITY GUIDE

**Completion Time**

☐ Read "Silicon Controlled Rectifiers." __________

☐ Answer Self Test Review Questions 1-10. __________

☐ Perform Experiment 9 on "SCR Characteristics." __________

☐ Read "Bidirectional Triode Thyristors." __________

☐ Answer Self Test Review Questions 11-19. __________

☐ Read "Bidirectional Trigger Diodes." __________

☐ Answer Self Test Review Questions 20-26. __________

☐ Read "Unijunction Transistors." __________

☐ Answer Self Test Review Questions 27-43. __________

☐ Perform Experiment 10 on "UJT Characteristics." __________

☐ Read the Unit Summary __________

☐ Complete the Unit Examination. __________

☐ Check the Examination Answers. __________

# SILICON CONTROLLED RECTIFIERS

The silicon controlled rectifier is the most popular member of the thyristor family. This device is generally referred to as an SCR. Unlike the bipolar transistor which has two junctions and provides amplification, the SCR has three junctions and is used as a switch. As its name implies, the device is basically a rectifier which conducts current in only one direction. However, the device can be made to conduct (turn on) or stop conducting (turn off) and therefore provide a switching action that can be used to control electrical current.

Let's now take a close look at the SCR. First you will examine its basic construction and operation, and then you will consider its important electrical characteristics and basic applications.

## Basic Construction and Operation

An SCR is a solid-state device which has four alternately doped semiconductor layers. The device is almost always made from silicon but germanium has been used. The SCR's four layers are often formed by a diffusion process but a combined diffusion-alloyed method is also used.

A simplified diagram of an SCR is shown in Figure 8-1A. As shown, the SCR's four (PNPN) layers are sandwiched together to form three junctions. However, leads are attached to only three of these layers. These three leads are referred to as the anode, cathode, and gate.

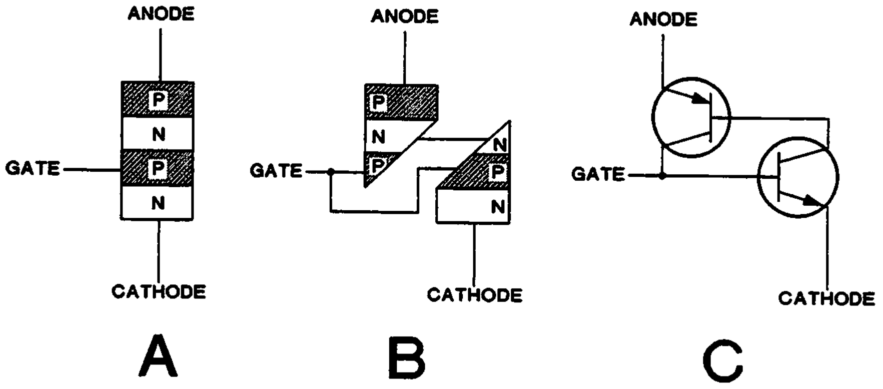


**Figure 8-1**
The SCR and its equivalent circuits.

The manner in which an SCR operates cannot be easily determined by examining the four-layer structure shown in Figure 8-1A. However, the SCR's four layers can be divided so that two three-layer devices are formed as shown in Figure 8-1B. The two devices obtained are effectively PNP and NPN transistors and when interconnected as shown they are equivalent to the four layer device shown in Figure 8-1A. These two transistors can also be represented by their schematic symbols as shown in Figure 8-1C.

We will now bias the equivalent circuit in Figure 8-1C just like we would bias the actual SCR shown in Figure 8-1A. First, we will make the anode of the circuit positive with respect to the cathode but we will leave the gate open. Under these conditions the NPN transistor will not conduct because its emitter junction will not be subjected to a forward bias voltage which can produce a base current. This will, in turn, cause the PNP transistor to turn off because the NPN transistor will not be conducting and therefore will not allow a base current to flow through the emitter junction of the PNP transistor. The equivalent SCR circuit will not allow current to flow from its cathode to its anode under these conditions.

If the gate of the equivalent SCR circuit is momentarily made positive with respect to the cathode, the emitter junction of the NPN transistor will become forward biased and this transistor will conduct. This will in turn cause a base current to flow through the PNP transistor which will cause this transistor to conduct. However, the collector current flowing through the PNP transistor now causes base current to flow through the NPN transistor. The two transistors therefore hold each other in the on or conducting state thus allowing current to flow continuously from the cathode to the anode of the circuit. It is important to note that this action takes place even though the gate voltage is applied only for a moment. This momentary gate voltage causes the circuit to switch to the on or conducting state and the circuit will remain in that state even though the gate voltage is removed.

In order to switch the equivalent SCR circuit back to its off or nonconducting state, it is necessary to reduce its anode-to-cathode voltage to almost zero. This will cause both transistors to turn off and remain off until the conditions previously described are repeated.

The SCR in Figure 8-1A operates just like the equivalent circuit in Figure 8-1C. In other words the SCR can be turned on by a positive input gate voltage and must be turned off by reducing its anode-to-cathode voltage. When the SCR is turned on and is conducting a high cathode-to-anode current, the device is said to be conducting in the forward direction. If the polarity of the cathode-to-anode bias voltage was reversed, the device would conduct only a small leakage current which would flow in the reverse direction.

The SCR is usually represented by the schematic symbol shown in Figure 8-2. Notice that this symbol is actually an ordinary diode symbol with a gate lead attached. The circle surrounding the diode may or may not be used and the SCR's anode, gate and cathode leads may or may not be identified. When the leads are identified, they are usually represented by the letters A, G, and K as shown.

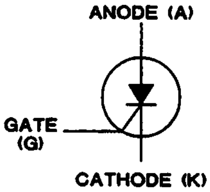


**Figure 8-2**
Schematic symbol of an SCR.

A properly biased SCR is shown schematically in Figure 8-3. Notice that a switch (S) is used to apply or remove the input gate voltage which is obtained from a voltage source and resistor Rg. This resistor is used to limit the gate current ($I_G$) to a specific value. The SCR's anode-to-cathode voltage is provided by another voltage source but a series load resistor ($R_L$) is also used to limit the SCR's cathode-to-anode current to a safe value when the device is turned on. Without the resistor, the SCR would conduct a very high cathode-to-anode current (also referred to as anode current or $I_A$) and could be permanently damaged.

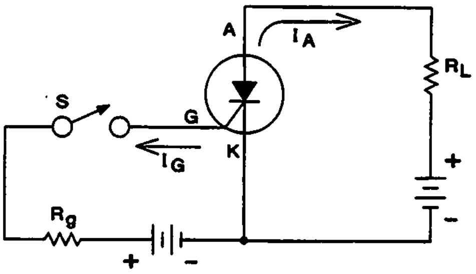


**Figure 8-3**
A properly biased SCR.

## V-I Characteristics

A better understanding of SCR operation can be obtained by examining the voltage-current (V-I) curve shown in Figure 8-4. This curve shows the V-I characteristics of a typical SCR. Such a curve is plotted by varying the SCR's cathode-to-anode voltage over a wide range while observing the SCR's anode current. The corresponding values are then plotted and a continuous curve is formed. The SCR is first biased in the forward direction while its gate is left open as shown in Figure 8-4. The SCR's cathode-to-anode voltage is designated as $V_F$

at this time. The curve shows that a $V_F$ increases from zero, the SCR conducts only a small forward current (designated as $I_F$) which is due to leakage. As $V_F$ continues to increase, $I_F$ remains very low and almost constant but eventually a point is reached where $I_F$ increases rapidly and $V_F$ drops to a low value (note the horizontal dotted line). The $V_F$ value required to trigger this sudden change is referred to as the forward breakover voltage. When this value of $V_F$ is reached the SCR simply breaks down, and conducts a high $I_F$ which is limited only by the external resistance in series with the device. The SCR switches from the off state to the on state at this time. The drop in $V_F$ occurs because the SCR's resistance drops to an extremely low value and most of the source voltage appears across the series resistor.

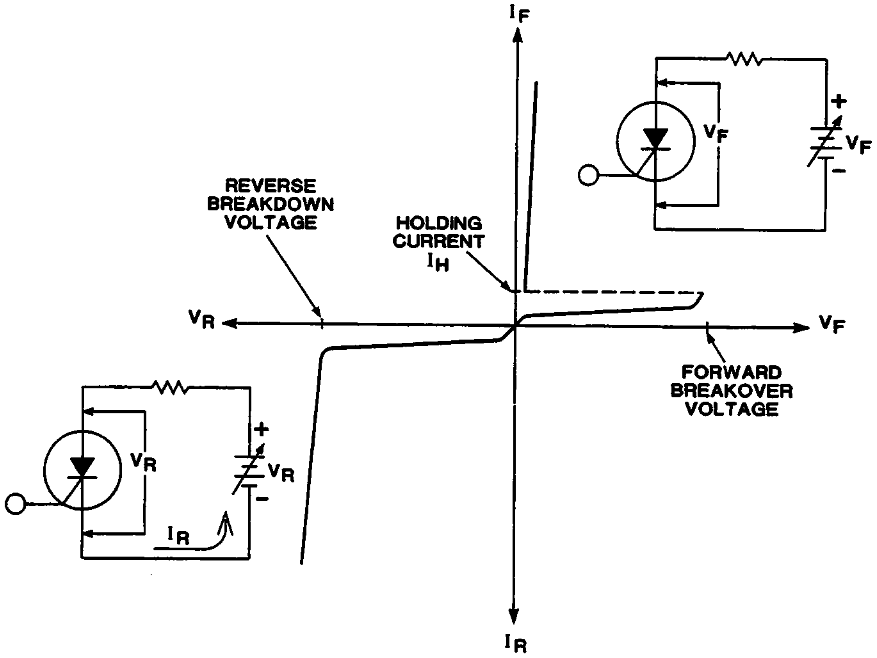


**Figure 8-4**
V-I characteristics of a typical SCR with the gate open.

When the SCR is in the on state, only a slight increase in $V_F$ is required to produce a tremendous increase in $I_F$ (the curve is almost vertical and straight). Furthermore, the SCR will remain in the on state as long as $I_F$ remains at a substantial value. Only when $I_F$ drops below a certain minimum value, will the SCR switch back to its offstate. This minimum value of $I_F$ which will hold the SCR in the on state is referred to as the SCR's holding current and is usually designated at $I_H$. As shown in Figure 8-4, the $I_H$ value is located at the point where breakover occurs (just to the left of the horizontal dotted line).

When a reverse voltage is applied to the SCR as shown in Figure 8-4, the device functions in basically the same manner as a reverse-biased PN junction diode. As the reverse voltage ($V_R$) across the SCR increases from zero, only a small reverse current ($I_R$) will flow through the device due to leakage. This current will remain small until $V_R$ becomes large enough to cause the SCR to breakdown. Then $I_R$ will increase rapidly if $V_R$ increases even slightly above the breakdown point (the curve is almost vertical and straight). The reverse voltage ($V_R$) required to breakdown the SCR is referred to as the SCR's reverse breakdown voltage. If too much reverse current is allowed to flow through the SCR after breakdown occurs, the device could be permanently damaged. However, this situation is normally avoided because the SCR is usually subjected to operating voltages which are well below its breakdown rating.

The V-I curve in Figure 8-4 shows the relationship between $V_F$ and $I_F$ when the SCR's gate is open. In other words no voltage is applied to the SCR's gate and therefore no gate current is flowing through the device. The curve in Figure 8-4 could therefore be labeled accordingly to indicate that gate current is equal to zero.

When the gate is made positive with respect to the cathode, gate current will flow and the SCR's forward characteristics will be affected. The changes that take place in the SCR's forward characteristics are graphically represented in Figure 8-5. In this figure, three V-I curves are plotted to show how changes in gate current (designated as $I_G$) affect the relationship between the SCR's forward voltage and forward current. The $I_G = 0.5$ curve shows the relationship between $V_F$ and $I_F$ when the gate current is zero and is therefore simply a more detailed representation of the forward characteristics shown in Figure 8-4. The $I_G = 1$ curve is plotted for a specific but relatively low value of gate current (more than zero). Notice that this curve has the same general shape as the $I_G = 0.5$ curve but the forward breakover point occurs sooner (at a lower $V_F$ value). The $I_G = 2$ curve is plotted for a slightly higher gate current and also has the same general shape as the other two curves. However, the breakover point occurs even sooner at this higher value of gate current.

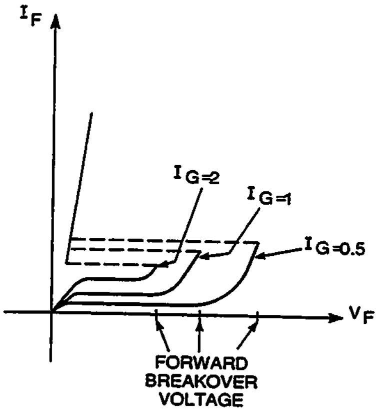


**Figure 8-5**
Forward characteristics of SCR
for different values of gate current.

The curves in Figure 8-5 show that the SCR's forward breakover voltage decreases as the gate current increases. In fact, the gate current could be increased to a point where the breakover voltage would be so low that the device would have characteristics that closely resemble those of an ordinary PN junction diode. This ability of the gate to control the point where breakover occurs is used to advantage in many types of electronic circuits.

The curves in Figure 8-5 reveal much about the SCR's most important electrical characteristics. Basically these curves show that for any given gate current, a specific forward breakover voltage must be reached before the SCR can turn on. However, the curves also show that for any given forward voltage across the SCR, a specific value of gate current must be reached before the device can turn on. Therefore, an SCR can be turned on only when it is subjected to the proper combination of gate current and forward voltage values.

# Practical Applications

In normal operation the SCR is subjected to forward voltages which are below its breakover voltage and the SCR is made to turn on by the application of a suitable gate current. This gate current is usually made high enough to insure that the SCR is switched to the on state at the proper time. Furthermore, the gate current is usually applied for just an instant in the form of a current pulse. A constant gate current is not required to trigger the SCR and would only cause more power to be dissipated within the device. Once the SCR turns on, it can be turned off only be reducing its forward current below its respective holding current value.

The SCR is primarily used to control the application of dc or ac power to various types of loads. It can be used as a switch to open or close a circuit or it can be used to vary the amount of power applied to a load. A very low current gate signal can control a very large load current. The SCR in Figure 8-3 is basically used as a switch to apply dc power to the load resistor ($R_L$) but in this basic circuit there is no effective means of turning off the SCR and removing power from the load. However this problem could be easily solved by simply connecting a switch across the SCR. This switch could be momentarily closed to short out the SCR and reduce its anode-to-cathode voltage to zero. This would reduce the SCR's forward current below the holding value and cause it to turn off.

A more practical SCR circuit is shown in Figure 8-6. In this circuit $SCR_1$ is used to control the dc power applied to load resistor $R_L$ and $SCR_2$ along with a capacitor (C) and a resistor ($R_1$) are used to turn off the circuit. When a momentary gate current flows through $SCR_1$, this SCR turns on and allows a dc

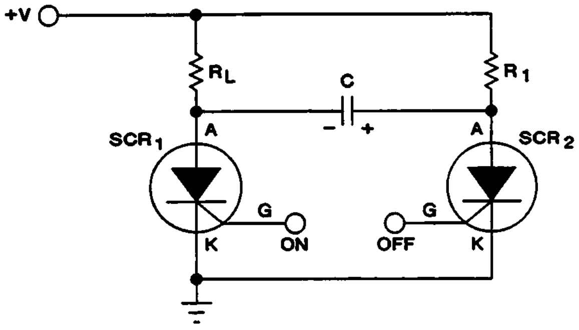


**Figure 8-6**
A practical dc SCR switch circuit.

voltage to be applied to $R_L$. This effectively grounds the left hand side of capacitor C and allows it to charge through resistor $R_1$. This, in turn, causes the right hand plate of C to become positive with respect to the left hand plate. When a momentary gate current pulse is applied to $SCR_2$, this SCR turns on and the right hand plate of the capacitor C is grounded thus placing this capacitor across $SCR_1$. The voltage across capacitor C now causes $SCR_1$ to be reverse-biased. This reverse voltage causes the forward current through $SCR_1$ to drop below its holding value thus causing this SCR to turn off and remove power from $R_L$. Therefore a momentary gate current through $SCR_1$ will turn on the circuit and a momentary gate current through $SCR_2$ will turn off the circuit.

When using SCR's in dc switching circuits it is often necessary to use additional components (not always additional SCR's) to provide a means of turning off the circuits. The previous example shows only one way this can be accomplished.

An SCR may also be used to control the application of ac power to a load. However, when used in ac circuits, the device is capable of operating on only one alternation of each ac input cycle. A simple ac switch circuit is shown in Figure 8-7. The SCR can conduct only on those alternations which make its anode positive with respect to its cathode. Furthermore, switch S must be closed so that gate current will flow through the SCR and allow it to conduct. Resistor $R_1$ limits the peak value of this gate current and diode $D_1$ prevents a reverse voltage from appearing between the SCR's gate and cathode during the reverse portion of each cycle. Therefore, by closing switch S the SCR is allowed to conduct on one alternation of each input cycle and apply this portion of the ac voltage to load resistor $R_1$. When switch S is opened, the SCR will run off within one-half cycle of the ac signal. In other words when no gate current can flow through the SCR, the device will turn off as soon as the input ac voltage drops to zero and starts increasing in the direction which causes the SCR's anode to be negative with respect to its cathode. The SCR will remain off until the switch is closed again.

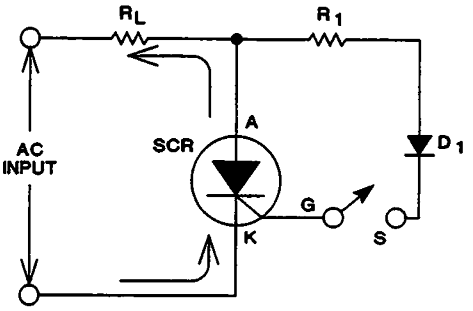


**Figure 8-7**

Simple ac SCR switch

The ac switch circuit in Figure 8-7 is actually less complicated than the dc circuit in Figure 8-6. In the ac circuit no additional components are required to turn off the SCR since this occurs automatically when the ac input voltage drops to zero. Instead, the additional components are used to insure that the SCR turns on during the proper portion of each cycle. It is also important to note that this switch circuit can deliver only half of the available ac power to the load since the SCR can conduct only on one alternation of each ac cycle. However, other circuit arrangements can be used which will allow full ac power to be applied to the load. This is often accomplished by first rectifying the ac input signal so that both alternations of each input cycle will properly bias the SCR into conduction. In other words the positive and negative alternations of each cycle are converted to all positive or all negative alternations and so that they become a pulsating dc current.

The circuits in Figure 8-6 and 8-7 are simply used to apply or remove electrical power and take the place of a mechanical switch. However, when compared to mechanical switches or relays, these SCR circuits have many advantages. They do not wear out like mechanical devices and they do not have contacts which can bounce or stick and cause intermittent operation. The SCR circuits are much more reliable than mechanical devices in applications where large amounts of power must be controlled. SCR circuits may in turn be controlled mechanically or electrically. In either case, it is only necessary to control the SCR's very small gate current. If this is done mechanically, a relatively inexpensive switch that has low current and voltage ratings can be used.

SCR's may also be used to vary the amount of power applied to a load instead of just simply switching the power on or off. In fact, they are widely used in power control applications where the source of power is 60 Hz ac. One of the most basic ac power control circuits which utilizes an SCR is shown in Figure 8-8. This circuit is commonly referred to as a half-wave phase control circuit. It uses just one SCR and it is capable of controlling the ac power applied to load resistor $R_L$.

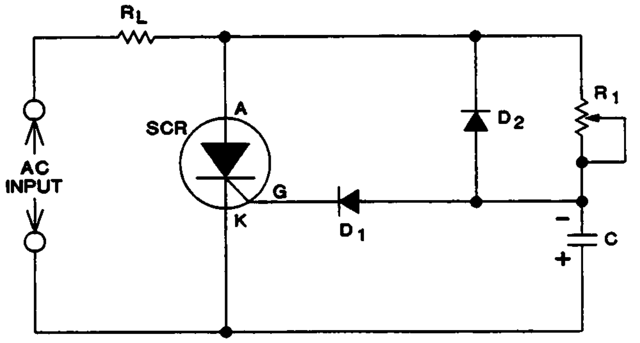


**Figure 8-8**
Half-wave phase control SCR circuit.

The ac voltage applied to the control circuit in Figure 8-8 is the standard 120 volts, 60 Hz sine wave. Two complete cycles of this waveform are show in Figure 8-9A. On each negative alternation of the input voltage, capacitor C charges through forward biased diode $D_2$ and the SCR is biased in the reverse direction so that it cannot conduct. Also, on each negative alternation, diode $D_1$ is reverse-biased and will not allow gate current to flow through the SCR. During each positive alternation the SCR is forward-biased so that it can conduct forward current through $R_L$ if its gate current is high enough to turn it on. However, the SCR's gate current is controlled by resistor $R_1$ and capacitor C. This capacitor will discharge and then recharge through $R_1$ during the positive alternations since $D_2$ is reverse-biased during these portions of the ac cycle. Furthermore, the rate at which capcitor C charges can be controlled by adjusting the resistance of $R_1$. If the resistance of $R_1$ is set to zero, capacitor C will charge almost immediately and the voltage across capacitor C (which is connected to the SCR's gate through $D_1$) will quickly rise to a level which will cause the SCR to turn on. In fact, when $R_1$ is equal to zero the SCR will turn on almost at the beginning of each positive alternation and apply power to $R_L$ for the entire alternation. Once the SCR turns on it will not turn off until the ac input voltage drops to zero. Therefore, when $R_1$ is equal to zero the SCR simple acts like an ordinary junction diode and conducts during each positive alternation of the input ac voltage. This causes the voltage across $R_L$ to appear as shown in Figure 8-9B.

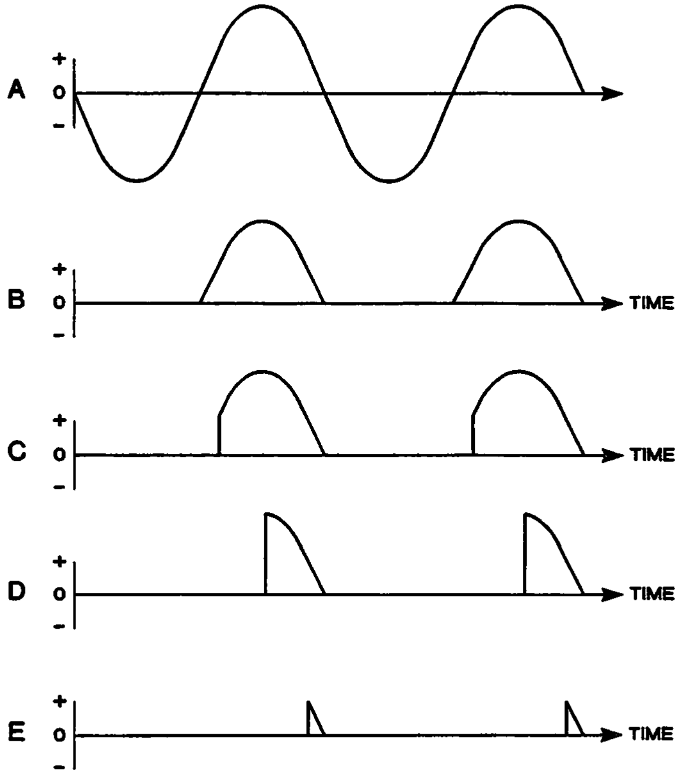


**Figure 8-9**
Waveforms for half-wave phase control SCR circuit.

When the resistance of $R_1$ is increased slightly, capacitor C cannot discharge and then recharge as quickly during each positive alternation. This means that it will take slightly longer for the voltage across capacitor C to rise to a level which will cause the SCR to turn on. However, when the SCR does turn on, it remains on for the remainder of each positive alternation. This means that the SCR turns on shortly after each positive alternation has started and not at the beginning of each alternation as it did before. This causes the voltage across $R_L$ to appear as shown in Figure 8-9C.

The resistance of $R_1$ can be further increased to further extend the time required for the SCR to turn on during each positive alternation. Figure 8-9D shows the voltage across $R_L$ when $R_1$ is increased to the point where the SCR conducts for only half of each positive alternation and Figure 8-9E shows the voltage across $R_L$ when $R_1$ is made even larger. If the value of $R_1$ is increased further, the SCR will not conduct at all and no power will be applied to $R_L$.

As shown in Figure 8-9, the half-wave phase control circuit is capable of controlling the amount of power applied to the load. With this circuit, the power applied to the load can be varied from zero to approximately 50 percent of the input ac power.

It is also possible to use an SCR to control current during both alternations of an ac signal, however it is necessary to first rectify the ac signal so that both alternations of each cycle are made to flow in the same direction (known as full wave rectification) before being applied to the SCR. Another method of achieving full control of an ac signal is to use two SCR's that are connected in parallel but in opposite directions. This arrangement allows each SCR to conduct alternately. Either of these two methods will allow full control of the input ac signal so that the power applied to the load can be varied from zero to the full value of the ac input.

## Self Test Review

1. An SCR has ____________ semiconductor layers.
2. The SCR has three leads which are referred to as the ______________, ____________, and ____________.
3. An SCR is used basically as an electronic ____________.
4. The SCR can be turned on or off and therefore has two stable states.
   a. True
   b. False
5. The forward voltage required to turn on an SCR is referred to as the SCR's ____________ ____________ ____________.
6. The minimum value of forward current required to keep an SCR in the on state is referred to as the SCR's ____________ ____________.
7. The reverse voltage that will cause an SCR's reverse current to change from an extremely small leakage value to a relatively high value is called the SCR's ____________ ____________ ____________.
8. An SCR's forward breakover voltage will decrease when its gate current is ____________.
9. Under normal conditions an SCR is biased so that it can be turned on with the application of a momentary ____________ current.
10. An SCR may be used to control the amount of ac power applied to a load.
    a. True
    b. False

## Answers

1. four
2. anode, cathode, and gate
3. switch
4. (a) True
5. forward breakover voltage
6. holding current
7. reverse breakdown voltage
8. increased
9. gate
10. (a) True

# EXPERIMENT 9

# SCR Characteristics

*OBJECTIVE:* *To demonstrate the operation of a typical SCR and show a practical application of the device.*

## Introduction

In this experiment, you will demonstrate the conditions required to turn on or turn off an SCR. You will do this by subjecting the SCR to its normal operating voltages and observing its operation through voltage measurements. Then, you will use a lamp to provide a visual indication of SCR operation. You will conclude the experiment by demonstrating the use of an SCR in a half-wave phase control circuit.

## Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
Oscilloscope
1 — Silicon Controlled Rectifier 2N5061 (57-624)
1 — Germanium diode (56-602)
1 — Silicon diode (56-28)
1 — 220 ohm, resistor (red, red, brown, gold)
1 — 1 kilohm, resistor (brown, black, red, gold)
1 — 4.7 kilohm, resistor (yellow, violet, red, gold)
1 — Lamp (412-601)
1 — 0.1 microfarad capacitor (21-786)

## Procedure

1. Turn on the Trainer and adjust the positive (+) voltage control for +10 volts. Turn the Trainer off.

2. Figure 8-10 is the schematic diagram for the circuit. Construct the circuit following the wiring diagram in Figure 8-11.

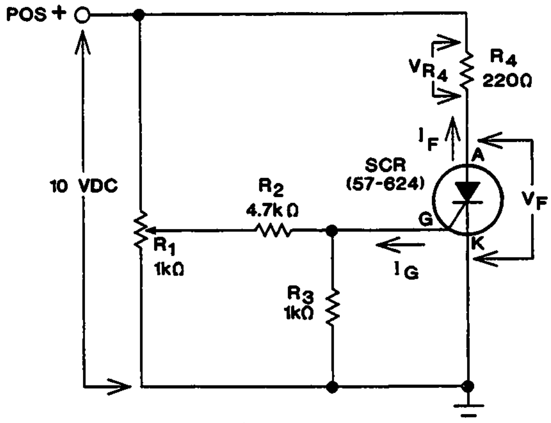


**Figure 8-10**
DC control SCR circuit for experiment 9.

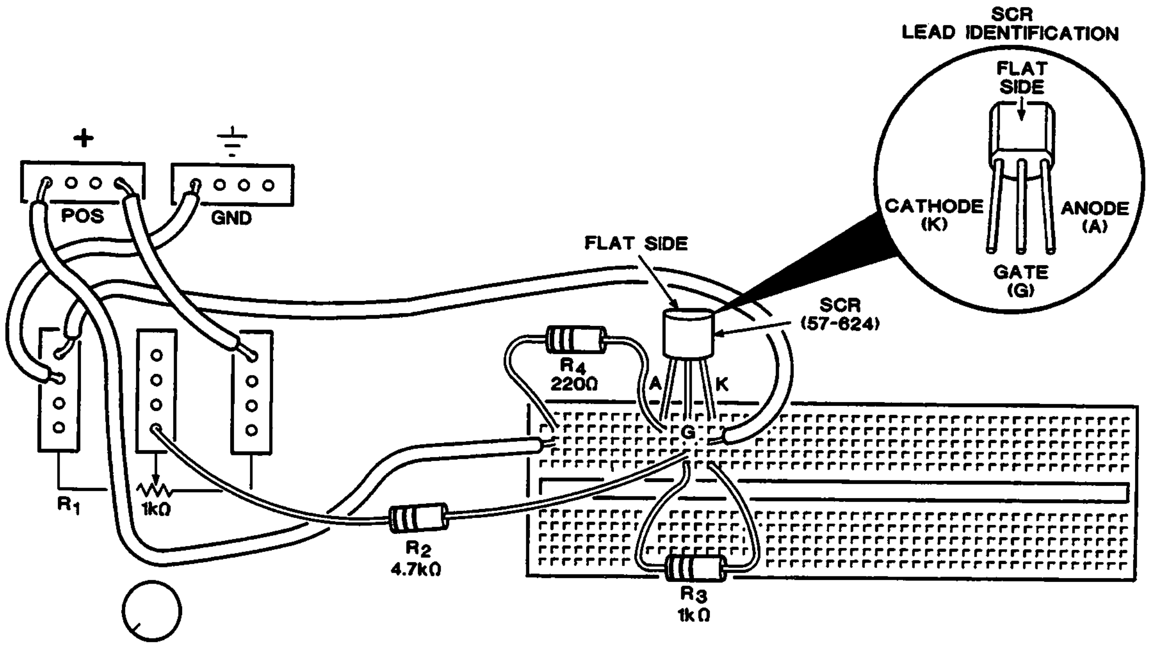


**Figure 8-11**
Wiring diagram for experimental circuit.

3. Turn the Trainer on and adjust the 1 kilohm potentiometer ($R_1$) for 0 volts at the wiper. $R_1$ is used to control the gate-to-cathode voltage, and the gate current ($I_G$). $I_G$ will be equal to zero with the present position of $R_1$.

4. Now measure the SCR's anode-to-cathode voltage (designated as $V_F$) and record this voltage reading in the appropriate space in the OFF column in the table provided in Figure 8-12. Your recorded value represents the forward voltage across the SCR when it is turned off.

| SCR CONDITION | OFF | ON |
|---|---|---|
| $V_F$(VOLTS) | | |
| $V_{R_4}$ (VOLTS) | | |
| $I_F$ (MILLIAMPERES) | | |

**Figure 8-12**

Table for measurements made in Steps 4 through 9.

5. Measure the voltage drop across $R_4$ and calculate the forward current ($I_F$) and record this value in the appropriate space in the off column in the table provided in Figure 8-12. This value represents the forward current ($I_F$) when the SCR is off.

6. Now use the voltage across $R_4$ and calculate the forward current ($I_F$) and record this value in the appropriate space in the off column in the table provided in Figure 8-12. This value represents the forward current ($I_F$) when the SCR is off.

7. Now connect your voltmeter across the SCR again and observe its forward voltage ($V_F$) as you adjust potentiometer $R_1$ slowly. This will cause the gate current ($I_G$) to increase. Continue to turn $R_1$ until you note a change in $V_F$. Then, stop turning $R_1$ and record the new $V_F$ value in the space provided in the ON column of Figure 8-12. Your recorded value represents the forward voltage across the SCR when it is turned on.

8. Measure the voltage across $R_4$ ($V_{R4}$) again and record this voltage in the space provided in the ON column in the table of Figure 8-12. Your recorded value represents the voltage across $R_4$ when the SCR is on.

9. Now use the $V_{R4}$ value obtained in step 8 and the resistance of $R_4$ to calculate the current through $R_4$. This calculated value represents the forward current ($I_F$) flowing through the SCR when it is turned on. Record this $I_F$ value (in milliamperes) in the space provided in the ON column in the table of Figure 8-12.

10. Now adjust $R_1$ to reduce $I_G$ to zero. Then, measure $V_F$ and $R_4$ and use these voltage readings to determine if the SCR is still on, or if the device has been turned off. According to these readings the SCR is in the ________ (on/off) state.

11. Without turning the Trainer off, disconnect one end of the $R_4$, and then, immediately install this resistor lead back in its original position. This will momentarily reduce $I_F$ to zero. Measure $V_F$ and $V_{R4}$ again, and determine if the SCR is on or off. According to these readings the SCR is now in the ____________ (on/off) state.

12. Turn off the Trainer. Then completely remove $R_4$, and in its place, install the lamp (412-601). This lamp will provide a visual indication of the SCR's operating state. The lamp will light (turn on) when the SCR switches to the "on" state. It will turn off when the SCR switches to the "off" state. Insure that $I_G$ equals zero, adjust potentiometer $R_1$ accordingly. Turn the Trainer on and observe the lamp. The lamp is now __________ (off/on).

13. Adjust potentiometer $R_1$ until the lamp changes state. The lamp is now __________ (off/on).

14. Adjust potentiometer $R_1$ to reduce $I_G$ to zero. The lamp is now __________ (off/on).

15. Turn off the Trainer. Wait for a few seconds and then turn the Trainer on again. The lamp is now __________ (off/on).

16. Turn off your Trainer.

## Discussion

In this portion of the experiment, you demonstrated the conditions necessary to turn on and turn off an SCR. First, you applied power to the SCR and its series resistor ($R_4$) but you did not allow gate current to flow through the device. Under these conditions the SCR was in the off state. Next, you measured the SCR's forward voltage ($V_F$) and the voltage across the series resistor ($V_{R4}$). You found that $V_F$ was almost equal to the applied voltage, or approximately 10 volts, and that $V_{R4}$ was approximately zero. You also calculated the SCR's forward current ($I_F$), and determined that it was equal to zero at that time.

Then, you increased the gate current ($I_G$) until the SCR turned on, and again measured $V_F$ and $V_{R4}$. You should have obtained a $V_F$ value of 1 volts or less and a corresponding $V_{R4}$ value of 9 volts or more. The lower $V_F$ value occurred because the SCR was conducting and its anode-to-cathode resistance was low. This allowed most of the supply voltage (10 volts) to appear across $R_4$. The supply voltage used in this experiment is relatively low, therefore, the voltage dropped across the SCR is noticeable. However, in many applications the supply voltage is very high (100 volts or more) and the voltage dropped across the SCR is insignificant compared to the voltage appearing across the load. While the SCR was turned on you also calculated its forward current ($I_F$). You probably obtained an $I_F$ value of 40 milliamperes or more.

Next, you reduced $I_G$ to zero and observed the condition of the SCR circuit by measuring $V_F$ and $V_{R4}$. $V_F$ should have remained low and unchanged, while $V_{R4}$ remained high and unchanged, thus indicating that the SCR remained in the on state.

Then, you momentarily disconnected one end of $R_4$ and observed the condition of the circuit by again measuring $V_F$ and $V_{R4}$. Your measurements should have indicated that the SCR was in the off state. The SCR switched to the off state because its forward current was reduced to zero, and was, therefore, forced to drop below the SCR's minimum holding current value. This holding current value was not actually determined since it is usually very low. The holding current for your SCR could be as low as several milliamperes.

Next, you replaced $R_4$ with a lamp. The lamp served as a series load to provide a visual indication of the SCR's operating condition. First, you applied power to the SCR circuit while holding $I_G$ to zero. The lamp should have remained off to indicate that the SCR was in the "off" state. Then, you increased $I_G$, and the lamp turned on to indicate that the SCR had changed state. Next, you adjusted $R_1$ to reduce $I_G$ to zero, however, the lamp should have remained on. this proved that once in the "on" state, reducing the gate voltage to zero would not turn off the SCR. Finally, you turned off your Trainer waited a few seconds and then turned the Trainer back on, while holding $I_G$ to zero. This should have caused the lamp to turn off, proving that in order to turn off the SCR $I_F$ must be momentarily reduced to zero by disconnecting the circuit or completely shutting it down. In either case, $I_F$ drops below the holding value and allows the device to unlatch (internally) and change state.

## Procedure

17. Construct the circuit shown in Figure 8-13. This potentiometer is used as a variable resistance in series with the 0.1 microfarad capacitor ($C_1$) to control the voltage at the gate of the SCR. The SCR, in turn, controls the ac voltage applied to the lamp. This circuit is identical to the half-wave phase control circuit described earlier; however, you will be using the 15 volt ac power provided on your Trainer instead of 120 volts ac.

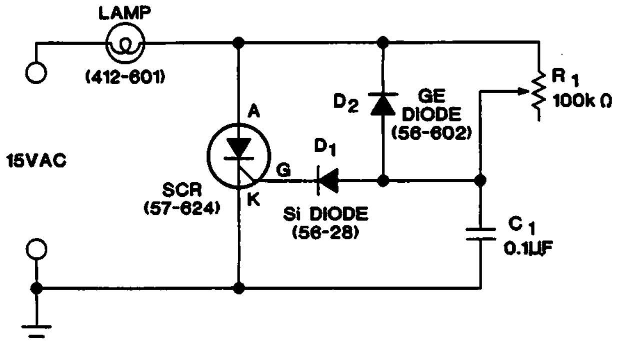


Figure 8-13A
ac phase control SCR circuit for experiment 8.

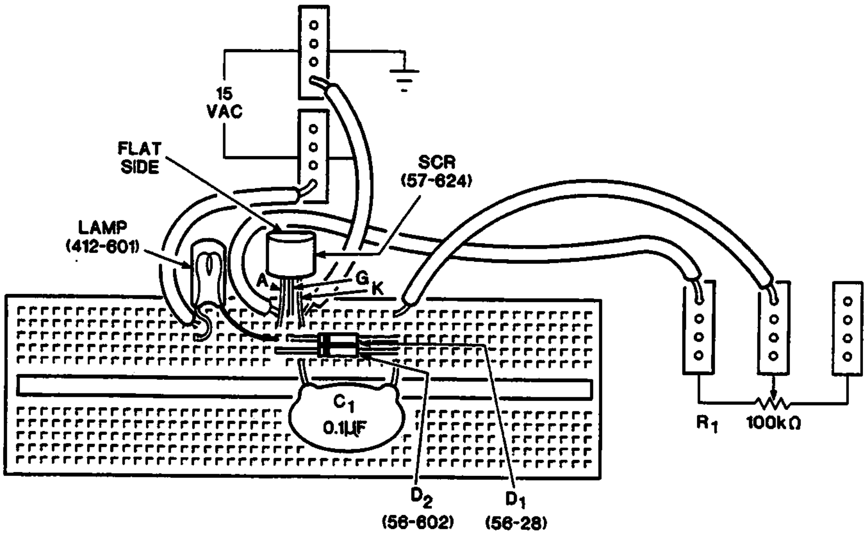


Figure 8-13B
Wiring diagram for experimental circuit.

18. Set the controls on your oscilloscope for 10 volts/cm and 5 milliseconds/cm. Connect the oscilloscope to the anode lead on the SCR. Adjust $R_1$ so that its resistance is minimum. Then turn on your Trainer.

19. Observe the lamp as you increase the resistance of $R_1$. Adjust $R_1$ until it is at maximum resistance. Then return $R_1$ to minimum resistance. Note the condition of the lamp as you adjust $R_1$. Also, observe the signal on the oscilloscope as you adjust $R_1$.

20. Turn off your Trainer.

## Discussion

In this part of the experiment, you demonstrated the use of an SCR in a half-wave phase control circuit. First, you adjusted $R_1$ for minimum resistance. This should have caused the lamp to glow at maximum intensity. Then you increased the resistance of $R_1$. This should have caused the intensity of the lamp to decrease. When $R_1$ was at maximum resistance, the lamp should have been completely off. Returning $R_1$ to minimum should have again increased the intensity of the lamp to maximum.

This experiment should have proved that the SCR can be used to vary the amount of power applied to a load. Although, ac power was applied to the input of the circuit, the SCR allowed only a pulsating dc current to flow through the load. You should have seen waveforms similar to those shown in Figure 8-9 as $R_1$ was varied. Furthermore, by adjusting $R_1$ the SCR's conduction time was varied, thus, controlling the average level of voltage applied to the load.

# BIDIRECTIONAL TRIODE THYRISTORS

The SCR described is capable of controlling current which is flowing in only one direction, and is therefore a unidirectional device. The SCR is used in many applications which involve the control of direct currents as well as alternating currents. Unfortunately when used in ac applications, a single SCR is capable of operating on just one alternation of each ac input cycle. In order to achieve full control of each ac input cycle, it is necessary to use two SCR's in parallel, or it is necessary to convert each ac cycle into a pulsating dc signal before it is applied to a single SCR.

In applications where it is necessary to achieve full control of an ac signal, it is often much easier to use a device known as a bidirectional triode thyristor. This device is more commonly referred to as a **triac**. The triac has basically the same switching characteristics as an SCR, however, it exhibits these same characteristics in both directions. This makes the triac equivalent to two SCRs in parallel, but connected in opposite directions.

We will briefly examine the triac's basic construction and operation. Then, we will consider its important characteristics and applications.

## Basic Construction and Operation

A simplified diagram of a triac is shown in Figure 8-14A. Notice that the device has three leads designated as main terminal 1, main terminal 2 and gate. Main terminal 1 and main terminal 2 are each connected to a PN junction at opposite ends of the device. The gate is also connected to a PN junction that is at the same end as terminal 1. If you examine the entire structure closely, you will see that from terminal 1 to terminal 2, you can pass through an NPNP series of layers or a PNPN series of layers. In other words, the triac is effectively a four-layer NPNP device in parallel with a four-layer PNPN device. These NPNP and PNPN devices are often compared to two SCRs which are connected in parallel, but in opposite directions. This equivalent SCR arrangement is shown in Figure 8-14B.

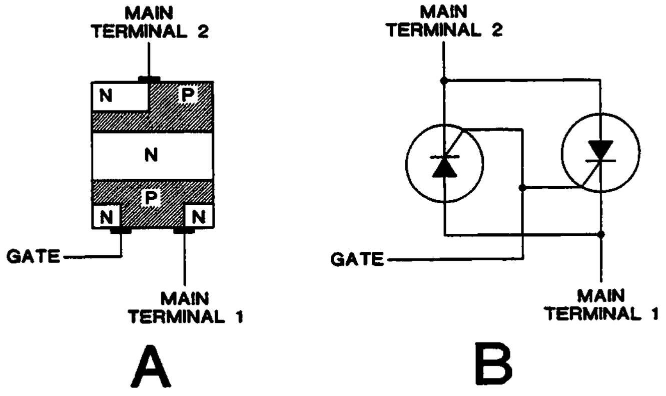


**Figure 8-14**

The triac and its equivalent circuit.

The triac's gate region is more complex and a detailed analysis of its operation will not be considered. However, the gate is basically capable of directly or remotely triggering either equivalent SCR into conduction. Notice that both of the equivalent SCR gates are tied together in Figure 8-14B to show the equivalent relationship.

The circuit in Figure 8-14B is not in all ways equivalent to the triac. This circuit is used simply to explain the basic concept involved. The primary difference is that the two equivalent SCRs would actually require different gating circuits to trigger them into conduction but the triac is designed to respond to the currents that flow through its single gate terminal.

Unlike an SCR which can control currents flowing in only one direction, the triac can control currents flowing in either direction. The triac is therefore widely used to control the application of ac power to various types of loads or circuits. The conditions required to turn a triac on or off in either direction are similar to the conditions required to control an SCR. Both devices can be triggered to the on state by a gate current and they can be turned off by reducing their operating currents below their respective holding values. In the case of an SCR, current must flow in the forward direction from cathode to anode. However, the triac is designed to conduct both forward and reverse currents through its main terminals.

The schematic symbol that is commonly used to represent the triac is shown in Figure 8-15. Notice that the symbol consists of two parallel diodes connected in opposite directions with a single gate lead attached. The device is usually placed within a circle as shown and its main terminals are sometimes identified as $MT_1$ and $MT_2$ as indicated.

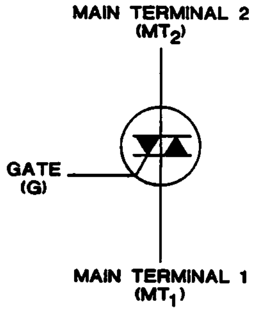


**Figure 8-15**
Schematic symbol of a triac.

# V-I Characteristics

The voltage-current (V-I) characteristic curve for a typical triac is shown in Figure 8-16. This curve shows the relationship between the current flowing through its main terminals in each direction (designated as $+I_T$ and $-I_T$) and the voltage applied across its main terminals in each direction (identified as +V and –V). Furthermore, this curve was plotted with no gate current flowing through the triac and main terminal 1 was used as the reference point for all voltage and current values.

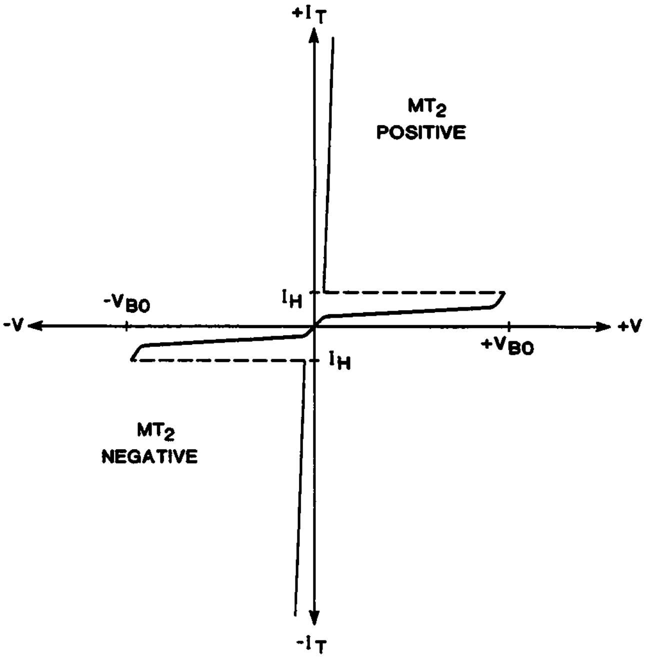


**Figure 8-16**
V-I characteristics of a typical triac.

Figure 8-16 shows that when main terminal 2 ($MT_2$) is positive with respect to $MT_1$ (or when the applied voltage is equal to +V), the current through the device ($+I_T$) remains at a low leakage value until +V rises above the breakover voltage ($+V_{BO}$) of the device. At this time the triac switches from the off state to the on state and $+I_T$ is essentially limited by the external resistance of the circuit. The triac must remain in the on state until $+I_T$ drops below a specified holding current ($I_H$) as shown. This is exactly what happens when an SCR is subjected to a forward voltage that exceeds its respective forward breakover voltage.

When $MT_2$ is negative with respect to $MT_1$, the triac exhibits the same basic V-I characteristics since the current through the device ($-I_T$) remains at a low leakage value until –V rises above the breakover voltage ($-V_{BO}$) of the device. At this time the triac switches from the off state to the on state and remains on until $-I_T$ drops below $I_H$.

The V-I curve in Figure 8-16 therefore, shows that the triac exhibits the characteristics of an SCR in either direction. However, this curve does not show how the triac's gate is used to control its operation. Like the SCR previously described, the triac's breakover voltage (in either direction) can be varied by controlling the amount of gate current flowing through the device. When the gate current is increased, the breakover voltage is lowered. However, there is still one very important difference between the SCR's gating characteristics and those of a triac. The SCR always requires a positive gate voltage, but the triac will respond to either a positive or negative gate voltage. In other words the triac's breakover voltage (in either direction) can be lowered by making its gate more positive or more negative with respect to $MT_1$ which is used as the reference terminal. This positive or negative gate voltage correspondingly produces a gate current that flows out of or into the gate lead and these currents in turn regulate the point at which the device turns on.

Like the SCR, the triac is normally subjected to operating voltages that are below its breakover voltage (in either direction). The device is turned on by subjecting it to a sufficiently high gate current which flows into or out of its gate lead. The device is turned off by simply reducing its operating current ($+I_T$ or $-I_T$) below its respective $I_H$ value. The triac is most sensitive when it is subjected to +V and $+I_T$ values along with a positive gate voltage. Under these conditions the device requires the least gate current to turn on for any given +V value. Other combinations of operating voltages and current result in somewhat less sensitivity. To help the circuit designer determine the conditions necessary to turn on a specific triac, manufacturers of these components usually specify minimum or typical values of gate current (in each direction) required to turn on the device. These values are given for a specified operating voltage that is applied in first one direction and then the other (+V and −V). With this information the circuit designer can insure that sufficient gate current is used to turn on the triac at the proper time. As with the SCR circuits previously described, this gate current need only be applied momentarily to cause the triac to change states.

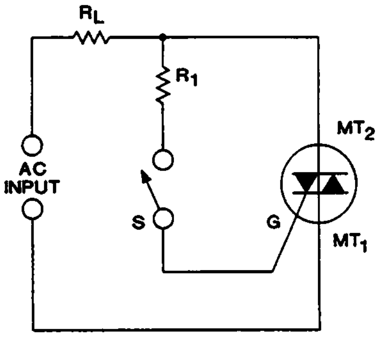


**Figure 8-17**
Simple ac switch.

# Applications

Since the triac conducts in either direction it is ideally suited for applications where ac power must be controlled. The device can be used as an ac switch or it can be used to actually control the amount of ac power applied to a load. A typical example of its use as an ac switch is shown in Figure 8-17. This circuit will apply the full ac input voltage across load resistor $R_L$ or completely remove it when switch S is closed or opened respectively. When switch S is open the triac cannot conduct on either the positive or the negative alternations of each ac input cycle since the input voltage does not exceed the triac's breakover voltage in either direction. However, when switch S is closed, resistor $R_1$ allows enough gate current to flow through the triac on each alternation to insure that the device turns on. The triac therefore applies all of the available input power to the load while a comparable SCR circuit (refer to Figure 8-7) is capable of supplying only half of the input power to the load. The advantage of this circuit is that the small gate current can control a high load current.

A typical triac circuit which can be used to vary the amount of ac power applied to a load is shown in Figure 8-18. This circuit is generally referred to as a full wave phase control circuit and it operates in a manner similar to the SCR circuit shown in Figure 8-8. The primary difference is that the triac is triggered into conduction on both the positive and negative alternations of each ac input cycle, while the SCR in Figure 8-8 conducts only on positive alternations. Also, a special triggering device is generally used to insure that the triac turns on at the proper time.

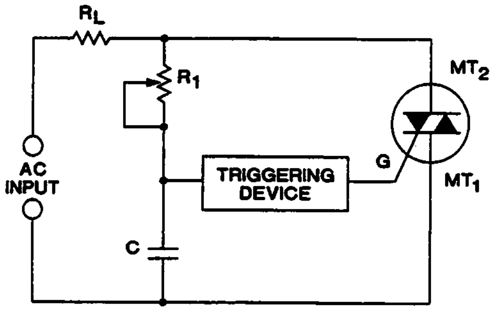


**Figure 8-18**
Full-wave phase control circuit.

Capacitor C charges through $R_1$ in first one direction and then the other as the positive and negative alternations of the ac input signal occur. During each alternation, the triac is turned on when the voltage across capacitor C rises to the required level. However, this voltage is not applied directly to the triac's gate and $Mt_1$ leads. Instead, it is applied through a special triggering device which has bidirectional switching characteristics. The triggering device could be any component which would turn on or conduct when subjected to a specific voltage level and turn off when the voltage is reduced to a lower level. One of the most widely used triggering devices will be described in detail later in this unit. However, at the present time we will just assume that the device is a solid-state component that has the switching characteristics just described.

During each alternation, the voltage across capacitor C rises to a level which turns on the triggering device. This causes current to momentarily flow through the triac's gate lead and switch it to the one state. The gate current only flows for a moment since the capacitor discharges through the triac and loses its accumulated voltage which in turn causes the triggering device to turn off. How soon the triac turns on during each alternation is determined by the value of $R_1$. When the resistance of $R_1$ is reduced to zero, the triac is triggered immediately at the beginning of each alternation and full ac power is applied to load resistor $R_L$. As the resistance of $R_1$ is increased, triggering occurs later during each alternation and the average power applied to the load is reduced. The voltage waveforms shown in Figure 8-9 can also represent the voltage across $R_L$ in the triac circuit if the negative alternations (which are identical but complementary to the positive alternations) are inserted.

A triggering device is required because the triac is not equally sensitive to gate currents flowing in opposite directions as explained earlier. The triggering device helps to compensate for the triac's non-symmetrical or non-uniform triggering characteristics. The voltage required to turn on the triggering device is identical in either direction and the device is designed to be as insensitive to temperature changes as possible. The triggering device works in conjuction with resistor $R_1$ and capacitor C to produce consistently accurate gate current pulses that are high enough to turn on the triac at the proper time in either direction. These gate current pulses can be very short in duration (several microseconds is generally sufficient) and still trigger the triac.

Although the triac has the ability to control current in either direction and respond to gate currents flowing in either direction, the device does have certain disadvantages when compared to an SCR. In general, triacs have lower current ratings than SCRs and cannot compete with SCRs in applications where extremely large currents must be controlled. Triacs are available that can handle currents (usually measured in rms values) as high as 25 amperes. By comparison, SCRs can be readily obtained with current ratings (usually expressed as average values for a half cycle) as high as 700 or 800 amperes, and some are rated even higher. Also, both devices can have peak or surge current ratings that are much higher than their respective rms or average ratings.

Triacs often have difficulty in switching the power applied to inductive loads. This problem also occurs with SCRs but to a lesser degree. When triacs are used to control the power applied to inductive loads such as motor windings or heater coils, it is always necessary to use additional components to improve their operation. Also, triacs are designed for low frequency (50 to 400 Hz) applications while SCRs can be used at frequencies up to 30 kHz. Therefore, in certain applications where full control of an ac signal is required, SCRs may operate more efficiently than triacs while in other applications the exact opposite may be true.

# Self Test Review

11. The triac is often compared to two____________ which are connected in parallel, but in opposite directions.

12. The triac's gate is capable of turning on the device so that it can conduct current in either direction through its $MT_1$ and $MT_2$ leads.

    a. True
    b. False

13. When the triac turns on in either direction, it can be turned off only by reducing its main operating current ($+I_T$ or $-I_t$) below a specified __________ current value.

14. The triac's breakover voltage, in either direction, is lowered when its gate current is _______________.

15. When a triac is subjected to an operating voltage that is less than its breakover voltage rating (in either direction) the device can be turned on by a gate current that is flowing in either direction.

    a. True
    b. False

16. A triac exhibits the same sensitivity to gate currents flowing in either direction.

    a. True
    b. False

17. When a triac is used to control the amount of ac power applied to a load, a special _________ device is used in conjunction with the triac to compensate for the triac's non-symmetrical triggering characteristics.

18. The triac in Figure 8-18 will apply full ac power to $R_L$ when $R_1$ is equal to __________.

19. Triacs should always be used instead of SCRs in applications where full control of ac power is required.

    a. True
    b. False

## Answers

11. SCR's
12. (a) True
13. holding
14. increased
15. (a) True
16. (b) False — The sensitivity of a triac varies for different gate current and operating voltage polarities.
17. triggering
18. zero
19. (b) False — In certain applications SCRs can provide more efficient control of ac power.

# BIDIRECTIONAL TRIGGER DIODES

As explained earlier, a triggering device is used in conjunction with the triac because the triac does not have symmetrical triggering characteristics. Various types of triggering devices can be used with the triac, but one of the most popular devices is known as a bidirectional trigger diode, commonly referred to as a *diac*.

We will now briefly examine the construction and operation of this important triggering device. Then we will see how it is used in conjunction with the triac.

## Basic Construction and Operation

The diac is constructed in much the same way as a bipolar transistor. The device has three alternately doped semiconductor layers as shown in Figure 8-19A. However, it differs from the bipolar transistor because the doping concentrations around both junctions are equal and leads are attached only to the outer layers. No electrical connections are made to its middle region. Since the diac has only two leads, it is often packaged in a metal or plastic case which has axial leads. Therefore the device often resembles an ordinary PN junction diode in appearance. However, the device is sometimes packaged like a conventional bipolar transistor but with only two leads.

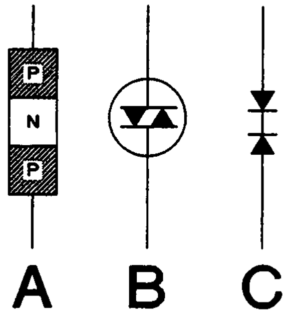


**Figure 8-19**
The diac, its schematic symbol and equivalent circuit.

Since both of its junctions are equally doped, the diac has the same effect on currents flowing in either direction through its leads. In either direction, one junction will always be forward-biased while the other is reverse-biased. In each case the reverse-biased junction primarily controls the current flowing through the diac and the device operates as if it contains two PN junction diodes that are in series but are connected back-to-back. See Figure 8-19C.

The diac remains in an off state (conducts only a small leakage current) in either direction until the applied voltage in either direction is high enough to cause its respective reverse-biased junction to break down. When this happens, the device turns on and current suddenly rises to a value which is essentially limited by the resistance in series with the device. The diac therefore functions as a bidirectional switch which will turn on whenever its breakdown voltage (in either direction) is exceeded.

A schematic symbol that is commonly used to represent the diac is shown in Figure 8-19B. Notice that it is similar to the triac symbol previously described, however no gate lead is required.

## V-I Characteristics

The voltage-current (V-I) characteristic curve for a typical diac is shown in Figure 8-20. This curve shows the relationship between the current flowing through the device in either direction (+I and -I) and the corresponding voltage across the device in each direction (+V and -V).

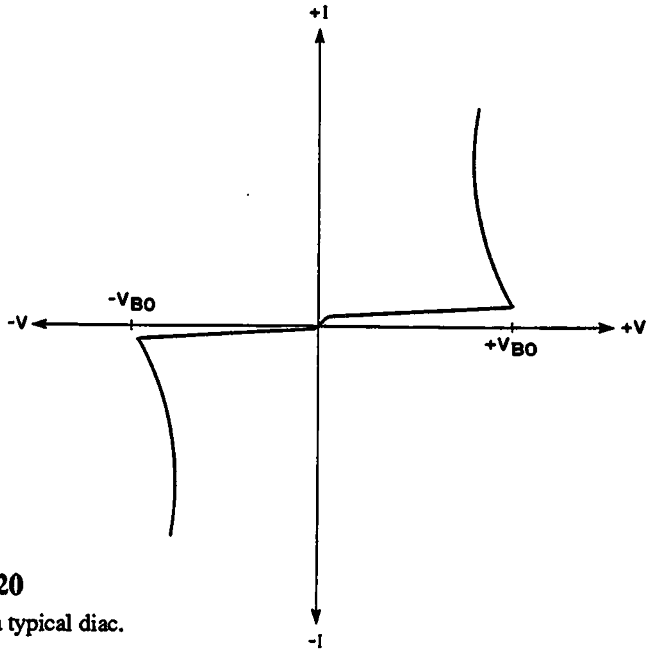


**Figure 8-20**
V-I Characteristics of a typical diac.

Figure 8-20 shows that the current through the diac remains at a low value until the voltage across the device increases to a point where the device breaks down in either direction. These voltages required to breakdown in the diac are generally referred to as the diac's breakover voltages and are designated as $+V_{BO}$ and $-V_{BO}$ as shown in Figure 8-20. The $+V_{BO}$ and $-V_{BO}$ values for a typical device will usually be between 28 and 36 volts. Notice that once the $+V_{BO}$ or $-V_{BO}$ values are reached, the current through the diac increases rapidly but the voltage across the device decreases. In other words when breakover occurs, the resistance of the diac decreases rapidly as current increases and the net result is a decrease in voltage across the device. Therefore, once the breakover point is reached the diac exhibits a negative resistance characteristic.

The diac is considered to be in the off state until the breakover voltage ($+V_{BO}$ or $-V_{BO}$) is reached. When $+V_{BO}$ or $-V_{BO}$ is reached, the device switches to the on state and is capable of conducting relatively high currents.

# Applications

The diac is most commonly used as a triggering device for a triac which is, in turn, used to control the amount of ac power applied to a load. This application is shown in Figure 8-21. The circuit shown in this figure is basically identical to the full-wave phase control circuit shown in Figure 8-18. The only difference is that a diac is specifically used as a triggering device. As explained earlier, the triggering device turns on when the voltage across $C_1$ rises to the required value. In this case the voltage across $C_1$ must rise to either $+V_{BO}$ or $-V_{BO}$ before the diac can switch from the off state to the on state. Each time the diac turns on, it allows current to momentarily flow through the triac's gate and turn it on. The gate current flows only momentarily because it is provided by capacitor $C_1$ which quickly discharges through the triac's $MT_1$ and gate terminals and the diac. The peak value and duration of the current pulse applied to the gate are determined by the value of $C_1$, the resistance of the diac (which changes with current), and the resistance between the triac's gate and $MT_1$ terminals. To help designers relate these various factors so that proper gate triggering will be obtained, manufacturers of diacs and triacs generally provide applicable charts or curves. The relationship between triggering capacitance (the value of $C_1$) and peak current for the diac is usually shown in this manner and various curves are used to show the minimum and maximum limits on the peak amplitude and duration of the gate current pulses which can be applied to a triac.

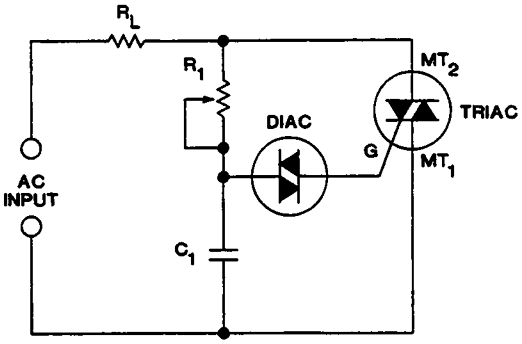


**Figure 8-21**
Full-wave phase control circuit using a diac and a triac.

The diac is used in conjunction with the triac to provide full-wave control of ac signals. These devices are commonly used together to control the speed and direction of electric motors or control the temperature of heating elements. They are used in air conditioning, heating, and ventilation systems and they can even be used in electronic garage-door systems. They may also be used to simply switch or control the amount of power applied to individual lamps or complete lighting systems. For example, they can be used to flash a lamp on and off as a warning signal, to control traffic-signal lights which are placed at roadway intersections to regulate traffic flow or provide light dimming in theaters. Diac-triac circuits can also be used to control power in three-phase systems by using multiple triac circuit arrangements.

Although the diac is widely used to trigger the triac in power control applications, it is important to realize that other triggering devices are also available. Most of these triggering devices have bidirectional switching characteristics that are closely related to those of a diac although each device is constructed differently and each is represented by a different schematic symbol. The diac just described serves as a typical example of a modern solid-state triggering device.

## Self Test Review

20. The diac has three semiconductor layers but only________ terminals.
21. The diac turns on when its ________ voltage is reached in either direction.
22. When the diac turns on in either direction, the device exhibits a ________ resistance.
23. When the diac is turned on, the voltage across the device decreases as the current through the device ________.
24. The diac is used in conjunction with a ________ to control the ac power applied to a load.
25. In Figure 8-21 the diac turns on when the voltage across______ rises to the diac's $+V_{BO}$ or $-V_{BO}$ values.
26. The diac is the only triggering device which can be used to control the triac.
    a. True
    b. False

### Answers

20. two
21. breakover
22. negative
23. increases
24. triac
25. capacitor $C_1$
26. (b) False — Other triggering devices are available with similar but not identical switching characteristics.

# UNIJUNCTION TRANSISTORS

We will now examine another important member of the thyristor family which is known as a unijunction transistor or simply a UJT. The UJT has physical and electrical characteristics that are quite different from those of the diacs and triacs previously described. The UJT is actually a special type of transistor which is used as an electronic switch and not as an amplifying device. Furthermore, the UJT can be used to generate repetitive waveforms which can be used to perform many useful functions in electronic circuits.

In addition to the basic unijunction transistor, a special type of UJT is also used. This device is commonly referred to as a programmed unijunction transistor or simply a PUT. The PUT operates in basically the same manner as the UJT previously mentioned but its electrical characteristics can be made to vary over a considerable range.

We will now examine the physical and electrical characteristics of the basic UJT and see how it is used as an electronic switch and waveform generator. Then we will briefly examine the special PUT and consider its advantages as compared to the UJT.

## Basic Construction

As its name implies, the unijunction transistor (UJT) has only one semiconductor junction. The device basically consists of a block or bar of N-type semiconductor material which has a small pellet of P-type material fused on its structure as shown in Figure 8-22A. A lead is attached to each end of the N-type bar as shown and these two leads are referred to as base 1 and base 2. Another lead is attached to the P-type pellet and referred to as the emitter. These leads are commonly designated $B_1$, $B_2$, and E, as shown.

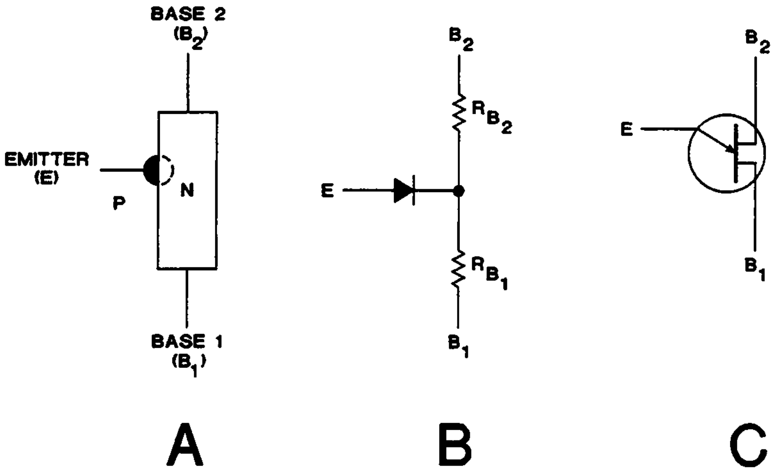


**Figure 8-22**

The basic UJT with its equivalent circuit and schematic symbol.

The N-type bar is lightly doped and therefore, has few majority carriers which can support current flow. This means that the resistance between the base 1 and base 2 leads is quite high. Most UJTs exhibit an interbase resistance (between $B_1$ and $B_2$) that is approximately 5 kilohms to 9 kilohms.

The area in which the P-type pellet and N-type bar meet to form the PN junction, has characteristics similar to those of a PN junction diode. Therefore, we can represent the UJT with the equivalent circuit shown in Figure 8-22B. This equivalent circuit consists of two resistors ($R_{B1}$ and $R_{B2}$) and a diode that has its cathode end connected between the resistors. The operation of the UJT is much easier to understand when this equivalent circuit is used in place of the UJT's basic structure as shown in Figure 8-22A.

The schematic symbol that is commonly used to represent the UJT is shown in Figure 8-22C. The arrow on the emitter (E) lead points inward to show that the UJT has a P-type emitter. The three leads may not always be identified as shown.

# Operation

The operation of an ordinary UJT is more apparent when its equivalent circuit is analyzed. Figure 8-23 shows how the equivalent UJT circuit is biased under normal conditions. Notice that an external voltage source ($V_{BB}$) is connected across the $B_1$ and $B_2$ terminals so that $B_2$ is positive with respect to $B_1$. Another external voltage source ($V_S$) is connected across the E and $B_1$ terminals in a manner which makes E positive with respect to $B_1$. A resistor is placed between the positive side of $V_S$ and terminal E. This resistor is used only to limit the current through E to a safe level.

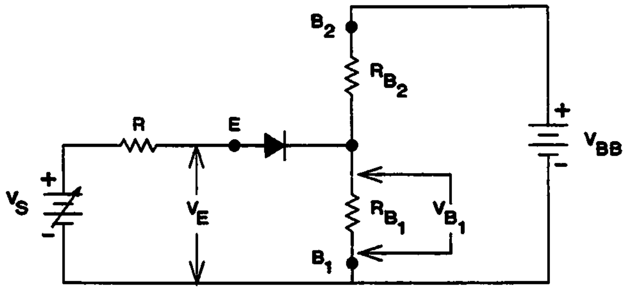


**Figure 8-23**
A properly biased UJT.

If voltage $V_S$ is not high enough to cause the diode in the equivalent circuit to be forward biased, the two resistors within the equivalent circuit ($R_{B1}$ and $R_{B2}$) will allow only a small current to flow between terminals $B_1$ and $B_2$. This current could be easily calculated according to Ohm's law by dividing the total resistance between terminals $B_1$ and $B_2$ ($R_{B1} + R_{B2}$) into voltage $V_{BB}$. This means that voltage $V_{BB}$ is distributed across $R_{B1}$ and $R_{B2}$. The ratio of the voltage across $R_{B1}$ (designated as $V_{B1}$) to the source voltage ($V_{BB}$) is known as the *intrinsic standoff ratio* and is represented by the Greek letter $\eta$ (eta). This relationship can be expressed mathematically as

$$\text{intrinsic standoff ratio } (\eta) = \frac{V_{B1}}{V_{BB}}$$

Since the voltage dropped across $R_{B1}$ and $R_{B2}$, is proportional to their resistance values, the intrinsic standoff ratio is also equal to the ratio of $R_{B1}$ to the total resistance between terminals $B_1$ and $B_2$ ($R_{B1} + R_{B2}$). This can be expressed mathematically as

$$\eta = \frac{R_{B1}}{R_{B1} + R_{B2}}$$

This last equation shows that the intrinsic standoff ratio is basically determined by the two internal resistances. Its value is, therefore, determined by the physical construction of the device and it cannot be controlled by varying $V_{BB}$ or $V_S$. The intrinsic standoff ratio is specified for each type of UJT that is made and typical values will range from approximately 0.5 to 0.8. When the intrinsic standoff ratio for a particular device is known, the voltage across $R_{B1}$ ($V_{B1}$) can be determined for any value of applied voltage ($V_{BB}$). This calculation can be made by simply transposing the first equation given above to obtain

$$V_{B1} = \eta\, V_{BB}$$

This equation simply states that the voltage across $R_{B1}$ ($V_{B1}$) is equal to the intrinsic standoff ratio times $V_{BB}$. For example, if a UJT has an intrinsic standoff ratio of 0.5 and it is subjected to a $V_{BB}$ of 20 volts, the voltage across $R_{B1}$ would be equal to 0.5 times 20 or 10 volts.

The UJT functions in the manner just described as long as $V_S$ is not high enough to forward-bias the diode and cause it to conduct. However, the UJT will exhibit considerably different characteristics when $V_S$ is high enough to forward bias the diode. In other words, the voltage across $R_{B1}$ ($V_{B1}$) causes the diode to be reverse-biased and this voltage must be completely canceled by the opposing input voltage ($V_E$). Thus, $V_E$ must be raised an additional 0.7 volts above $V_{B1}$ so that the diode will become forward-biased by a voltage that is high enough to cause it to conduct.

The value of $V_E$ required to turn on the diode is called the *peak voltage* and is usually designated as $V_P$. The value of $V_P$ is determined by the source voltage $V_{BB}$, the intrinsic standoff ratio, and the voltage required to turn on the diode. This relationship can be shown mathematically as

$$\text{peak voltage } (V_P) = \eta\, V_{BB} + V_F$$

This equation simply states that $V_P$ is equal to the product of the intrinsic standoff ratio and source voltage plus the voltage required to turn on the diode ($V_F$). Manufacturers usually do not indicate the value of $V_P$ for a particular UJT since it varies with $V_{BB}$. For example, if a UJT has an $\eta$ of 0.6 and a $V_{BB}$ of 10 volts, its peak voltage will be

$$V_P = (0.6)\,(10) + 0.7$$

$$V_P = 6 + 0.7$$

$$V_P = 6.7 \text{ volts}$$

This means that the diode will turn on when input voltage $V_E$ reaches 6.7 volts.

Until the $V_P$ value is reached, the diode conducts only a very small reverse leakage current which flows through the emitter lead. However, when the $V_P$ value is reached, the diode turns on and allows current to flow in the forward direction through its PN junction and through the emitter lead. This current (often designated as $I_E$) results because the diode turns on, but an additional action also takes place. When $I_E$ begins to flow in the forward direction, many charge carriers (holes) are injected into the lightly doped N-type bar and are swept toward terminal $B_1$ which is negative with respect to the emitter. These charge carriers increase the conductivity of the bar between the emitter (E) and $B_1$ which of course means that the resistance of $R_{B1}$ is reduced. This lower resistance causes $I_E$ to increase even further and more charge carriers are injected into the bar which in turn lowers the value of $R_{B1}$ even further. This in turn causes $I_E$ to increase even further. This action is cumulative and it starts to occur when the $V_P$ value is reached. At this time the UJT is said to be turned on or triggered to its on state.

If the source voltage ($V_S$) is increased further, the cumulative action just described becomes even more apparent. As $V_S$ increases, $I_E$ increases rapidly due to the decrease in the resistance of $R_{B1}$. Furthermore, this decrease in the value of $R_{B1}$ causes the voltage $V_E$, which appears between the E and $B_1$ terminals, to decrease even though $I_E$ increases in value. The UJT therefore exhibits a negative resistance characteristic after it turns on.

If $V_S$ is increased even further, $I_E$ would continue to increase, however a point would eventually be reached where $V_E$ would stop decreasing and actually begin to rise slightly. This point marks the end of the negative resistance region. Beyond this point an increase in $I_E$ is accompanied by a slight increase in $V_E$.

So far we have considered only the action that takes place between terminals E and $B_1$ when the UJT turns on. We have seen that $R_{B1}$ decreases, however it is important to note that $R_{B2}$ also decreases a certain amount since some of the injected charge carriers (holes) enter that portion of the N-type bar. However, the decrease in the value of $R_{B2}$ is small compared to the decrease in $R_{B1}$. This means that the total resistance of the bar ($R_{B1}$ + $R_{B2}$) is reduced once the UJT turns on and this results in an increase in current through these two resistors and voltage source $V_{BB}$. However, $R_{B2}$ does not decrease by much and therefore does not allow the current to rise to a high value.

The most important action takes place between terminals E and $B_1$. It is this portion of the UJT that provides the most useful characteristics. The two most important features are the ability to turn on at a specific $V_P$ value and the negative resistance characteristic that occurs for a certain period of time after the device turns on.

## V-I Characteristics

The action that takes place between the E and $B_1$ terminals of an ordinary UJT is shown graphically in Figure 8-24. The curve in this figure shows the relationship between the current flowing through the emitter of a typical UJT (designated as $I_E$) and the voltage appearing across its E and $B_1$ terminals (designated as $V_E$). Such a curve could be plotted by using a circuit like the one shown in Figure 8-23. The $V_E$ and $I_E$ values could be observed while $V_S$ is varied over a wide range.

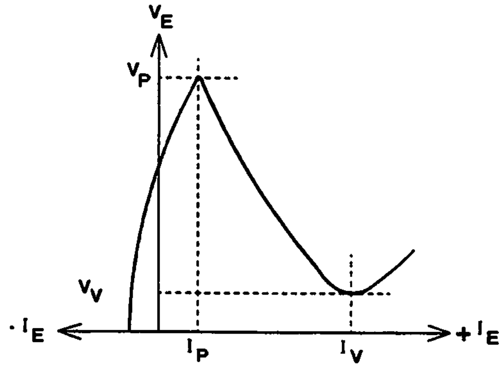


**Figure 8-24**
V-I characteristic between
E and $B_1$, terminals of a typical UJT.

Figure 8-24 shows that when $V_E$ equals zero a small negative current ($-I_E$) flows through the emitter lead. This is a small leakage current which flows from left to right through the diode (refer to Figure 8-23) because of the relatively large voltage across $R_{B1}$ ($V_{B1}$). As $V_E$ increases, it opposes $V_{B1}$, and the leakage current decreases. When $V_E$ is equal to $V_{B1}$ the current is reduced to zero and any further increase in $V_E$ results in a positive current ($+I_E$) which flows from right to left through the diode. When $V_E$ reaches the peak voltage ($V_P$) value, the UJT is considered to be in the on state. The current that flows at this point is called the *peak current* and is designated as $I_P$ as shown in Figure 8-24.

Beyond the $V_P$ point, $V_E$ decreases as $+I_E$ increases thus giving the device a negative resistance characteristic. This negative resistance continues until $V_E$ starts to increase again. The point where $V_E$ reaches its minimum value and starts to increase is called the *valley voltage* and is designated as $V_V$. The current flowing at this time is referred to as the *valley current* or $I_V$. Beyond the $V_V$ point, $V_E$ increases slightly as $+I_E$ increases and the UJT no longer exhibits a negative resistance.

## Applications of the UJT

The UJT's negative resistance characteristic makes the device useful for generating repetitive signals. A circuit that is commonly used for this purpose is shown in Figure 8-25. This circuit is commonly referred to as a UJT relaxation oscillator and it is capable of generating two types of waveforms which can be used in a variety of applications.

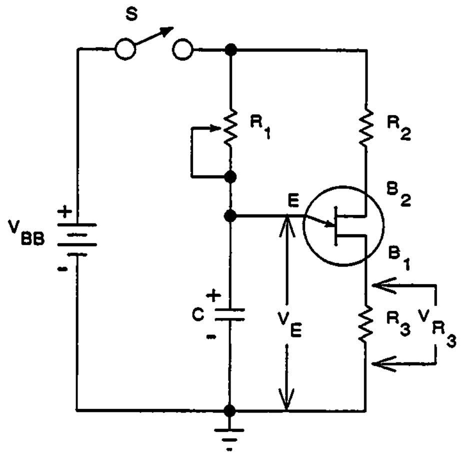


**Figure 8-25**
UJT relaxation oscillator.

When switch S is closed, capacitor C charges through resistor $R_1$ in the direction shown. When the voltage across C reaches the UJT's $V_P$ value. the UJT turns on and the resistance between terminals E and $B_1$ decreases. This allows C to discharge through the UJT and resistor $R_3$ (which has a very low value). The voltage across C quickly decreases to the UJT's $V_V$ value and this causes the UJT to stop conducting or turn off. As soon as the UJT turns off, capacitor C starts charging again and continues until $V_P$ is again reached. This causes the UJT to turn on which in turn allows C to discharge until $V_V$ is reached. This action continues with the voltage across C rising slowly to $V_P$ and decreasing rapidly to $V_V$. The voltage appearing between the UJT's E and $B_1$ terminals, fluctuates as shown in Figure 8-26A. Notice that this voltage follows a sawtooth type of pattern and after the initial turn-on it varies only between $V_V$ and $V_P$.

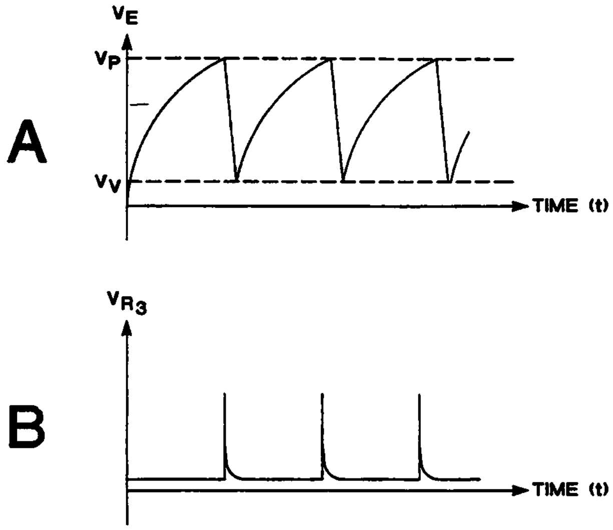


**Figure 8-26**
UJT relaxation oscillator waveforms with respect to ground at the emitter (A) and $B_1$ (B) terminals.

Each time capacitor C discharges current is momentarily forced through $R_3$. These momentary pluses of current cause the voltage across $R_3$ ($V_{R3}$) to pulsate accordingly. These momentary voltage pulses are shown in Figure 8-26B. Notice that the pulses are very narrow or sharp and a pulse occurs each time C discharges. Also notice that the voltage never drops completely to zero. A small voltage will always be dropped across $R_3$ because a small current flows through the UJT (from $B_1$ to $B_2$) even when the device is in the off state.

The number of complete sawtooth waveforms or pulses produced each second (frequency of oscillation) can be controlled by adjusting $R_1$. If the resistance of $R_1$ decreases, capacitor C will charge faster and the circuit will operate at a higher frequency. When $R_1$ increases in value, it takes longer for C to charge to $V_P$ and the frequency decreases. The frequency can also be varied by replacing capacitor C with a larger or smaller unit. A larger capacitor would charge slower while a smaller capacitor would charge at a faster rate.

The sawtooth and pulse waveforms produced by the UJT relaxation oscillator can be used to perform various functions in electronic circuits. However, the UJT is also used in applications where it is not required to oscillate continuously. For example, the UJT can function as a bistable (two state) element and switch from its on state to its off state or vice-versa when it receives an appropriate input signal. The device can also be used as a frequency divider. Another important application is its use as a triggering device. The UJT is capable of producing current pulses which are ideally suited for triggering SCRs. In fact, the UJT could be used to trigger the SCR in the half-wave phase control circuit shown in Figure 8-8. In other words, the voltage across capacitor C in Figure 8-8 could be used to trigger the UJT and the UJT could in turn generate the trigger pulses needed to turn on the SCR at the proper time.

## Programmable UJTs

The programmable UJT or PUT is one of the newest members of the thyristor family. This device is not constructed like an ordinary UJT (it uses four semiconductor layers), however it is capable of performing basically the same functions as the UJT. Also, the PUT has three terminals or leads which are referred to as the *cathode*, *anode* and *gate*.

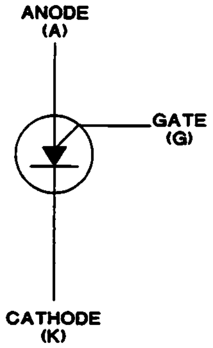


**Figure 8-27**
Schematic symbol for a PUT.

The primary difference between the ordinary UJT and the PUT is the fact that the peak voltage ($V_P$) of the PUT can be controlled. In the UJT, the $V_P$ value is fixed and cannot be externally controlled. The PUT's anode and cathode terminals are used in much the same way as the UJT's E and $B_1$ terminals and the anode is always made positive with respect to the cathode. In fact, the V-I characteristics between the PUT's anode and cathode terminals are identical to the V-I characteristics between the UJT's E and $B_1$ terminals. Both devices exhibit the same turn on characteristics and both have a negative resistance region between their peak and valley points. Furthermore, the $V_P$ value of the PUT is controlled by varying the voltage between the gate and cathode terminals with the gate always at a positive potential with respect to the cathode.

The operation of the PUT can be clearly demonstrated when it is connected in a circuit like the one shown in Figure 8-28. This circuit is referred to as a PUT relaxation oscillator and it produces the same basic waveforms shown in Figure 8-26. Notice that the gate-to-cathode voltage (called the gate voltage) is obtained from resistor $R_4$ which is part of a voltage divider network consisting of $R_3$ and $R_4$. This gate voltage is not appreciably affected by resistor $R_2$. This resistor has a very low value and is used to develop the output voltage pulses. As long as the gate voltage remains constant, the PUT will remain in its off or nonconducting state until its anode-to-cathode voltage (called its anode voltage) exceeds the gate voltage by an amount equal to the voltage drop across a single diode (approximately 0.7 volts). At this time the $V_P$ value is reached and the device turns on.

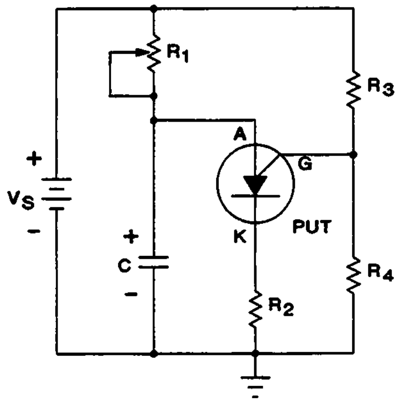


**Figure 8-28**
PUT relaxation oscillator.

The anode voltage is obtained from capacitor C which charges toward the supply voltage ($V_S$) through $R_1$. This is similar to the action that takes place in Figure 8-25. Each time the voltage across C reaches the $V_P$ value, the PUT turns on and allows C to discharge through its anode and cathode terminals and $R_2$. When the voltage across C drops to the $V_V$ value of the PUT, the device turns off and the cycle is repeated. A sawtooth voltage is therefore developed across C and voltage pulses appear across $R_2$ which resemble the waveforms shown in Figure 8-26. As before, the frequency of oscillation can be varied by adjusting $R_1$, or by changing the value of C. However, in this circuit the frequency can also be varied by changing the ratio of $R_3$ and $R_4$. It is this resistance ratio that controls the PUT's gate voltage which in turn controls the value of $V_P$. If $R_4$ is made larger while $R_3$ remains constant, the gate voltage will increase and this will cause $V_P$ to increase. The higher $V_P$ value will make it necessary for capacitor C to charge to a higher voltage before the PUT can turn on This will increase the time required to generate each sawtooth or pulse and decrease the frequency of operation. If $R_3$ increases in value while $R_4$ remains constant, the action will be exactly opposite and frequency will increase.

The frequency of the PUT relaxation oscillator in Figure 8-28 can be controlled by adjusting the ratio of $R_3$ and $R_4$ to control the PUT's $V_P$ value. Although this is an extremely important consideration, it is also important to realize that resistors $R_3$ and $R_4$ control other PUT characteristics as well. The peak current ($I_P$) and valley current ($I_V$) both depend on the values of $R_3$ and $R_4$ as well as the value of the source voltage ($V_S$). For example, it is possible to adjust $R_3$ and $R_4$ for any specific ratio so that any specific $V_P$ is obtained (taking into account the value of $V_S$). However, it is only the ratio of $R_3$ and $R_4$ that determine $V_P$, not the individual values of these resistors. The actual resistor values determine the PUT's $I_P$ and $I_V$ values for any given value of $V_S$.

This relationship between the external voltage divider network ($R_3$ and $R_4$) and the values of $I_P$ and $I_V$ is somewhat complex. However, this relationship can be simplified by relating $I_P$ and $I_V$ to the combined values of $R_3$ and $R_4$. $R_3$ and $R_4$ are considered to present an equivalent resistance ($R_G$) to the gate of the PUT which is equal to their parallel equivalent:

$$R_G = \frac{R_3 R_4}{R_3 + R_4}$$

For any given value of $V_S$, $I_P$ and $I_V$ both will decrease as $R_G$ increases. More complex equations may also be used to determine the exact values of $I_P$ and $I_V$ under various circuit conditions or these values can be determined graphically when appropriate charts or curves are provided. The circuit designer may use either of these methods to calculate the required values and in general all of the quantities just described must be considered when designing an oscillator circuit like the one shown in Figure 8-28. The intent here is not to show a complete design procedure for a PUT oscillator; but to emphasize the PUT's important electrical characteristics.

In addition to its use in relaxation oscillators, the PUT may also be used as a triggering device for SCRs. Due to its four-layer construction the PUT is capable of supplying trigger pulses that have a higher amplitude than those obtained from ordinary UJTs and can therefore trigger SCRs that have higher current ratings. The PUT is also used in long duration timer circuits (which are basically identical to the circuit in Figure 8-28) because the device has a low gate-to-anode leakage current. This is important in timer circuits since this leakage current adds to the charging current flowing through the capacitor. In such a circuit the capacitor must charge over a long period of time before the $V_P$ value is reached and an additional charging current (leakage) cannot be tolerated.

The PUT's important electrical characteristics can be controlled, thus making it more versatile than an ordinary UJT. The PUT is also more sensitive than a UJT and it responds faster. It is therefore used in place of the UJT in a number of applications although the UJT is still preferred in certain types of circuits.

# Self Test Review

27. The UJT has one semiconductor junction.
    a. True
    b. False
28. The leads attached to each end of the UJT's N-type bar are referred to as ________ and ________.
29. The lead that is attached to the UJT's P-type pellet is referred to as the ________.
30. A PUT has ________ PN junctions.
31. An equivalent UJT circuit can be formed with two resistors ($R_{B1}$ and $R_{B2}$) and a ________.
32. The ratio of $R_{B1}$ to the total resistance ($R_{B1}$ + $R_{B2}$) is called the ________ ________ ________.
33. The input voltage value ($V_E$) required to turn on a UJT is referred to as the ________ ________ and is designated as $V_P$.
34. The current that flows through the UJT's emitter lead at the point where $V_P$ is reached is referred to as the ________ ________ and is designated as $I_P$.
35. Beyond the $V_P$ point $V_E$ decreases until it reaches a minimum value which is referred to as the ________ ________ and is designated as $V_V$.
36. The current that flows through the UJT's emitter lead at the $V_V$ point is referred to as the ________ ________ and is designated as $I_V$.
37. Between the $V_P$ and $V_V$ points the UJT exhibits a ________ resistance.
38. When used as a relaxation oscillator, the UJT can generate both sawtooth and pulse type waveforms.
    a. True
    b. False
39. The UJT is often used as a triggering device for the ______.
40. The PUT can perform the same basic functions as a UJT although it is constructed differently and has three leads which are referred to as the ________, ________ and ________.
41. The PUT's $V_P$ value can be controlled by changing the voltage between its ________ and ________.
42. In Figure 8-28 the PUT's $V_P$ value is varied by changing the ratio of ________ and ________.
43. If $R_3$ increases in value while $R_4$ remains constant, the circuit in Figure 8-28 will operate at a ________ frequency.

## Answers

27. (a) true
28. base 1 and base 2
29. emitter
30. three (it is a 4 layer device)
31. diode
32. intrinsic standoff ratio
33. peak voltage
34. peak current
35. valley voltage
36. valley current
37. negative
38. (a) true
39. SCR
40. anode, cathode, and gate
41. gate and cathode
42. $R_3$ and $R_4$
43. higher

# EXPERIMENT 10

# UJT Characteristics

*OBJECTIVE:* *To demonstrate the operation of a typical unijunction transistor and show a practical application of the device.*

## Introduction

In the first part of this experiment, you will determine the peak voltage ($V_P$) and valley voltage ($V_V$) values for a typical UJT, and you will use the $V_P$ value that you obtained to determine the UJT's intrinsic stand off ratio. In the second part of this experiment, you will demonstrate the operation of a UJT relaxation oscillator.

## Material Required

Heathkit Analog Trainer
Voltmeter
Oscilloscope
1 — Unijunction transistor 2N4894 (417-183)
1 — 47 ohm, resistor (yellow, violet, black, gold)
1 — 100 ohm, resistor (brown, black, brown, gold)
1 — 1 kilohm, resistor (brown, black, red, gold)
1 — 10 kilohm, resistor (brown, black, orange, gold)
1 — 330 kilohm, resistor (orange, orange, yellow, gold)
1 — 10 microfarad electrolytic capacitor (25-880)

## Procedure

1. Plug in your Trainer, but be sure that the unit is turned off.

2. Construct the circuit shown in Figure 8-29 following the wiring diagram shown in Figure 8-30. The leads on the UJT (417-183) are identified in Figure 8-30. Turn the + voltage control fully counterclockwise (minimum voltage) and then adjust the 1 kilohm potentiometer (designated as $R_1$) for maximum resistance.

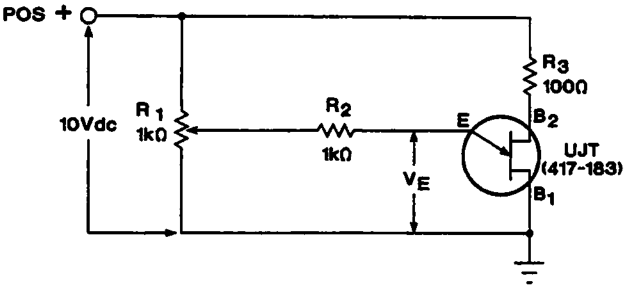


**Figure 8-29**
UJT circuit for experiment 10.

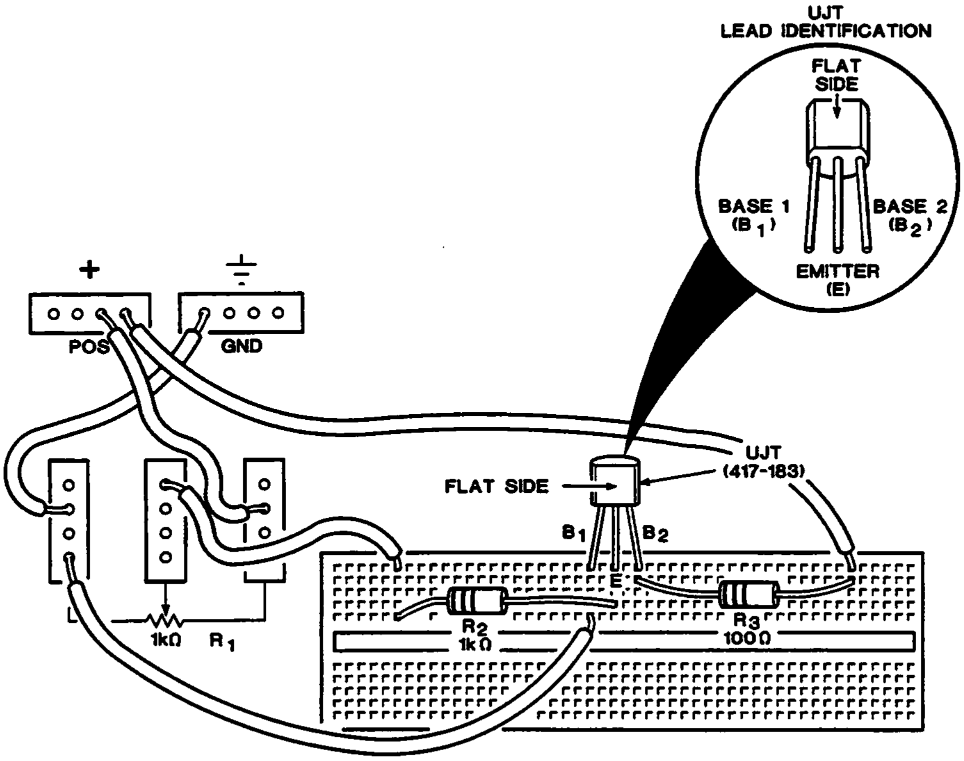


**Figure 8-30**
Wiring diagram for experimental circuit.

3. Turn on your Trainer and adjust the + voltage control for 10 volts at the output of the power supply. Potentiometer $R_1$ can now be used to control the UJT's emitter-to-base 1 input voltage ($V_E$) with the 1 kilohm resistor ($R_2$) serving as the series emitter resistor. The 100 ohm resistor ($R_3$) is used in the circuit to prevent the base 1-to-base 2 current from exceeding a safe value and does not significantly affect the normal operation of the circuit.

4. Connect the leads of your voltmeter between the UJT's emitter and base 1 terminals to observe the input voltage ($V_E$). Gradually turn $R_1$ clockwise as you observe $V_E$ and continue to turn $R_1$ until $V_E$ reaches a peak value and decreases. Note the peak or maximum $V_E$ value that you obtained. This maximum value of $V_E$ represents the UJT's peak voltage ($V_P$). Since $V_E$ drops suddenly after the $V_P$ value is reached, you may fail to detect the peak value the first time. In this case, adjust $R_1$ for maximum resistance, and start again. Record the $V_P$ value that you obtained in the space provided below.

$V_P$ = ________________

5. Continue to observe voltage $V_E$ as you rotate $R_1$ to maximum resistance. Notice that once the $V_P$ value is exceeded, voltage $V_E$ rapidly drops to a low value and then increases only a small amount. You will find that it is necessary to observe $V_E$ very closely to detect this slight increase. The $V_E$ value that you are observing (with $R_1$) is approximately equal to the UJT's valley voltage ($V_V$). Record this $V_V$ value in the space provided below.

$V_V$ = __________

6. Now you will use the $V_P$ value that you obtained in step 4 to determine the UJT's intrinsic standoff ratio ($\eta$). You will use the equation for $V_P$ ($V_P = \eta V_{BB} + V_F$) which can be rearranged to show that the intrinsic standoff ratio is equal to:

$$\eta = \frac{V_P - V_F}{V_{BB}}$$

When using this equation remember that $V_F$ is the forward voltage drop across the UJT's internal PN junction or approximately 0.7 volts and that $V_{BB}$ is the voltage source which is 10 volts. Record your calculated value of $\eta$ in the space provided below.

$\eta$ = _________

7. Now adjust $R_1$ to maximum resistance. Then, adjust the + voltage control for +5 volts at the output of the power supply.

8. Repeat steps 4 and 5, and record the $V_P$ and $V_V$ values that you obtain in the spaces provided below.

$V_P$ = ____________________

$V_V$ = ____________________

9. Calculate the UJT's intrinsic standoff ratio using the $V_P$ value obtained in step 8 and the $V_{BB}$ value of 5 volts, and record your results below.

$\eta$ = ____________________

10. Turn off your Trainer.

## Discussion

In this portion of the experiment, you measured the UJT's $V_P$ and $V_V$ values while the voltage source was equal to 10 volts. You should have obtained a $V_P$ value that was approximately 8 volts. The $V_V$ value that you obtained should have been equal to 2 volts, or less, which is typical for most UJT's.

Next, you used the $V_P$ value to calculate the UJT's intrinsic standoff ratio. You probably obtained a ratio that was between 0.7 and 0.9 however, a slightly higher or lower ratio may have resulted. Your UJT is designed to have a minimum ratio of 0.74 and a maximum ratio of 0.86, but minor errors in your voltage readings plus the assumption that $V_F$ is always equal to 0.7 volts could cause the calculated ratio to be somewhat different than the true value.

Next, you adjusted the voltage source to 5 volts and again measured $V_P$ and $V_V$. You should have discovered that the decrease in source voltage ($V_{BB}$) from 10 volts to 5 volts (a ratio of 2 to 1) caused the UJT's $V_p$ value to decrease by an almost proportional amount. The decrease cannot be exactly proportional because of the constant voltage (0.7 volts) that is dropped across the UJT's internal PN junction. Therefore, your new $V_P$ value was probably in the neighborhood of 4 volts or almost one-half of its previous value. Your new $V_V$ value may have decreased slightly but no significant change should have been noted. The $V_F$ value tends to remain relatively constant in a UJT even when its $V_{BB}$ value is varied over a considerable range.

You completed this portion of the experiment by sing the new $V_P$ value and the new $V_{BB}$ value (5 volts) to calculate the UJT's intrinsic standoff ratio. You should have obtained approximately the same ratio, thus, proving that it remains essentially constant even though the operating voltage is changed. The intrinsic standoff ratio of any UJT has a fixed value that is determined by the internal resistances within the device. If you noted any change, it was probably due to errors in voltage measurements.

## Procedure (continued)

11. Construct the circuit shown in Figure 8-31. Use any wiring arrangement that you choose but be sure to connect the UJT's leads as shown. Also be sure to connect the positive end of the 10 microfarad electrolytic capacitor to the UJT's emitter lead as shown. This circuit is basically identical to the UJT relaxation oscillator circuit shown in Figure 8-25. However, resistor $R_1$ has a fixed value and cannot be used to adjust the frequency of the circuit. The frequency is fixed at a very low value so that one complete sawtooth or pulse waveform is produced approximately every 5 seconds.

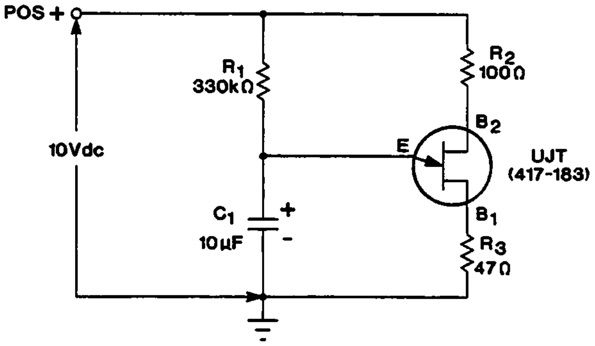


**Figure 8-31**
UJT relaxation oscillator circuit for experiment 10.

12. Turn on your Trainer and adjust the + voltage control until +10 volts is applied to the circuit. Set the oscilloscope controls, volt/cm for 2 volts and Time/cm for 10 milliseconds.

13. Measure the voltage across $C_1$ with your oscilloscope. This voltage will gradually rise to a maximum value and then rapidly drop to a minimum value to produce a sawtooth waveform, as shown in Figure 8-26A. Note the maximum/minimum voltages and record them below.

maximum voltage = ________________
minimum voltage = ________________

14. Monitor the voltage across $R_3$ with your oscilloscope. This voltage will occur in pulses, as shown in Figure 8-26B. Observe this voltage for at least 10 to 15 seconds so that you will see at least two voltage fluctuations.

15. Turn off your Trainer. Remove the 330 kilohm resistor ($R_1$) and install a 10 kilohm resistor. Then turn on your Trainer.

16. Measure the voltage across $C_1$ with your oscilloscope. Observe the maximum and minimum voltage values and also note the frequency of the voltage fluctuations. Record the values below and indicate if the frequency is higher or lower than it was in step 13.

maximum voltage = ________________
minimum voltage = ________________
Frequency is ____________

17. Turn off your Trainer.

## Discussion

In this portion of the experiment, you used your oscilloscope to observe the operation of a UJT relaxation oscillator. The signal you observed should have resembled the signal shown in Figure 8-26A. The maximum voltage was equal to the UJT peak voltage ($V_P$). The $V_P$ value was determined in step 4. Each time this $V_P$ value is reached the UJT turns on and the voltage across $C_1$ immediately drops to the UJT's valley voltage ($V_V$) value and causes the device to turn off.

When you monitored the voltage across $R_3$ you should have observed voltage pulses or fluctuations that were low in amplitude. A voltage pulse occurred each time the UJT turned on and allowed $C_1$ to discharge through $R_3$.

When you replaced the 330 kilohm resistor ($R_1$) with a smaller 10 kilohm resistor and observed the voltage across $C_1$. you should have obtained a maximum value less than $V_p$ and a minimum value considerably more than the $V_V$ value. Thus, indicating that the frequency is much higher. Your results should prove that the frequency can be increased by decreasing the value $R_1$ or vice-versa.

# UNIT SUMMARY

The SCR is the most popular member of the thyristor family. This device has four alternately doped semiconductor layers and three leads which are referred to as the anode, cathode, and gate. This device acts like an electronic switch which can be turned on by a momentary gate voltage and turned off by reducing its operating current below a certain holding value. The SCR may be used to switch dc or ac power and it can even be used to vary the amount of ac power applied to a load.

The triac effectively performs the same function as two SCRs that are connected in parallel but in opposite directions. It can therefore control ac currents more efficiently than a single SCR since it operates on both the positive and negative portions of each ac input cycle. The triac must be used in conjunction with a triggering device because of its nonsymmetrical triggering characteristics.

The triggering device most commonly used with the triac is known as a diac. The diac is simply a bidirectional solid-state switch which turns on in either direction when its breakdown voltage in each direction is exceeded. The diac has symmetrical switching characteristics since its breakdown voltage is approximately the same in each direction. When it is used to trigger a triac, it simply forces the triac to turn on each time it turns on.

Although it has characteristics that are different from those of SCRs, triacs, and diacs, the unijunction transistor (UJT) is still considered to be a thyristor. This device has only one PN junction and three leads (base 1, base 2, and the emitter). The UJT's important electrical characteristics occur between its emitter and base 1 leads. The device will turn on and conduct a substantial current through these leads when the voltage across these leads reaches a certain maximum value known as the peak voltage ($V_P$). Beyond this point the device exhibits a negative resistance until a certain minimum voltage known as the valley voltage ($V_V$) is reached. These characteristics allow the UJT to generate repetitive signals when used in a relaxation oscillator circuit or it can be used as a bistable switching element or as a triggering device.

A special type of UJT known as a programmable UJT or PUT is also available. This device operates like an ordinary UJT but its $V_P$ value as well as other important electrical characteristics can be varied by an external control voltage. This makes the PUT more versatile than the UJT and therefore suitable for a broader range of applications. The PUT is widely used as a triggering device and it is often used in long duration timer circuits.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understand of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows:

1. An SCR will switch to the on state only when its

   A. gate current is zero.
   B. forward breakover voltage is exceeded.
   C. reverse breakdown voltage is exceeded.
   D. gate current is below a certain threshold value.

2. An SCR will switch to the off state when its

   A. gate current is reduced.
   B. gate current is increased.
   C. forward current is increased to a sufficiently high value.
   D. forward current drops below a level which will sustain conduction.

3. An SCR is normally biased so that it can be turned on by

   A. momentarily increasing its forward voltage.
   B. increasing its reverse voltage.
   C. momentarily applying a gate current.
   D. momentarily decreasing its gate current.

4. When used in a half-wave phase control circuit an SCR is capable of varying the power applied to a load over a range that extends from zero to

   A. 50 percent of the ac input power.
   B. 100 percent of the ac input power.
   C. 75 percent of the ac input power.
   D. 90 percent of the ac input power.

5. A triac can control current flowing in either direction through its

   A. $MT_2$ and gate terminals.
   B. $MT_1$ and $MT_2$ terminals.
   C. $MT_1$ and gate terminals.
   D. cathode and anode terminals.

6. The triac can be turned on in either direction by a

   A. positive gate voltage but not a negative gate voltage.
   B. gate current that flows out of the gate terminal only.
   C. gate current that flows into the gate terminal only.
   D. sufficiently high gate current that flows in either direction.

7. The triac will switch to the off state when its main operating current (in either direction) drops below a specified value known as its

A. breakdown current.
B. threshold current.
C. holding current.
D. leakage current.

8. When a triac is used in a full-wave phase control circuit, it can vary the power applied to a load over a range that extends from zero to

A. 50 percent of the ac input power.
B. 100 percent of the ac input power.
C. 75 percent of the ac input power.
D. 90 percent of the ac input power.

9. A device that is commonly used to trigger the triac is referred to as a

A. bipolar transistor.
B. triggering thyristor.
C. SCR.
D. diac.

10. The UJT turns on and exhibits a negative resistance between its __________ and __________ terminals after its $V_P$ value is reached.

A. emitter and base 1.
B. emitter and base 2.
C. base 1 and base 2.
D. anode and cathode.

11. The UJT's negative resistance region is located between the

A. $V_P$ and $I_P$ pints on its V-I curve.
B. $V_P$ and $V_V$ points on its V-I curve.
C. $V_V$ and $I_V$ points on its V-I curve.
D. $+V_P$ and $-V_P$ points on its V-I curve.

12. The PUT has the same V-I characteristics as an ordinary UJT, however its $V_P$ value can be varied by changing its

A. gate-to-cathode voltage.
B. gate-to-anode voltage.
C. anode-to-cathode voltage.
D. anode-to-cathode current.

# EXAMINATION ANSWERS

1. B — The forward breakover voltage is that value of forward bias voltage which will cause the SCR to breakdown and switch to the on state. The breakover voltage can be adjusted by varying the SCR's gate current.

2. D — This minimum sustaining current is commonly referred to as the SCR's holding current.

3. C — The gate current is usually applied for only an instant and its amplitude is high enough to insure that the SCR turns on.

4. A — Since the SCR can only conduct during one-half of each input cycle, the maximum power applied to the load cannot exceed 50 percent of the ac input power.

5. B — The current through the $MT_1$ and $MT_2$ terminals can be switched on or off in either direction by subjecting the triac to the proper operating currents and voltages.

6. D — The triac will respond to gate current in either direction, however, its sensitivity will not be exactly the same in each direction.

7. C — The triac can be turned off only by reducing its operating current below its respective holding value. Furthermore, it is necessary to reduce the current for only a moment.

8. B — The triac can apply approximately the full ac input power to the load.

9. D — The diac is also known as a bidirectional trigger diode.

10. A — The negative resistance characteristic occurs only between the emitter and base 1 terminals.

11. B — Between these points voltage decreases as current increases thus producing a negative resistance characteristic.

12. A — The PUT's gate-to-cathode voltage controls its $V_P$ value and therefore controls the point at which the device turns on.

## *Unit 9*

# INTEGRATED CIRCUITS

# CONTENTS

Introduction . . . . Page 9-3

Unit Objectives . . . . Page 9-4

Unit Activity Guide . . . . Page 9-5

The Importance of Integrated Circuits . . . . Page 9-6

  Advantages . . . . Page 9-7

  Disadvantages . . . . Page 9-9

IC Construction . . . . Page 9-12

  Monolithic IC's . . . . Page 9-12

    Bipolar IC's . . . . Page 9-14

    MOS IC's . . . . Page 9-19

  Thin-Film IC's . . . . Page 9-21

  Thick-Film IC's . . . . Page 9-23

  Hybrid IC's . . . . Page 9-24

  IC Packages . . . . Page 9-25

Applications of IC's . . . . Page 9-34

  Digital IC's . . . . Page 9-34

  Linear IC's . . . . Page 9-41

Experiment 11 . . . . Page 9-44

Unit Summary . . . . Page 9-49

Unit Examination . . . . Page 9-51

Examination Answers . . . . Page 9-53

# INTRODUCTION

In this unit you will examine a relatively new type of solid-state device that is known as an *integrated circuit* or IC. The integrated circuit is actually a group of extremely small solid-state components which have been formed within or on a piece of semiconductor material and then appropriately interconnected to form a complete circuit. The IC is therefore a solid-state circuit and not an individual solid-state component like a diode or a transistor.

Integrated circuits are constructed in basically four different ways. They can be produced as *monolithic* devices, as *thin-film* or *thick-film* devices, or as *hybrid* devices. IC's may also be divided into two major groups according to their mode of operation. They are designed for either *linear* operation or *digital* operation.

The integrated circuit is an extremely important solid-state device which is now widely used throughout the electronics industry. You must be familiar with this important device if you plan to service, design, or intimately work with almost any type of electronic equipment. This unit explains why integrated circuits are used, how they are constructed, and where they find their greatest applications. Study this unit carefully, as it contains information that will help you throughout your electronics career.

# UNIT OBJECTIVES

When you have completed this unit on integrated circuits you should be able to:

1. State the need for integrated circuits in electronics.
2. Name at least three advantages that integrated circuits have over conventional circuits.
3. Name three disadvantages of using integrated circuits as compared to conventional circuits.
4. Explain the difference between monolithic, film-type, and hybrid integrated circuits.
5. Explain the difference between linear and digital integrated circuits.
6. Recognize three basic integrated circuit packages.
7. Explain the difference between SSI, MSI, and LSI circuits.

# UNIT ACTIVITY GUIDE

|  | Completion Time |
|---|---|
| ☐ Read "The Importance of Integrated Circuits." | ________ |
| ☐ Answer Self Test Review Questions 1-10. | ________ |
| ☐ Read "IC Construction." | ________ |
| ☐ Answer Self Test Review Questions 11-30. | ________ |
| ☐ Read "Applications of IC's." | ________ |
| ☐ Answer Self Test Review Questions 31-43. | ________ |
| ☐ Perform Experiment 11 on "Operational Amplifier Characteristics." | ________ |
| ☐ Read the Unit Summary. | ________ |
| ☐ Complete the Unit Examination. | ________ |
| ☐ Check the Examination Answers. | ________ |

# THE IMPORTANCE OF INTEGRATED CIRCUITS

Integrated circuits, or IC's, have changed the entire electronics industry. Before IC's were developed, all electronic circuits consisted of individual, or discrete, components that were wired together, often requiring a large amount of physical space. Printed circuit board technology made it possible to reduce the amount of space required. However, electronic circuits can be quite complex, requiring a large number of components, and since discrete components have a fixed physical size, there is a practical limitation on the amount of size reduction that can be achieved. Therefore, an important consideration in the design of circuits containing discrete components is the amount of space required by the circuit. However, the development of integrated circuit technology has made it possible to fabricate large numbers of electronic components on a single silicon chip. As a result, the physical size of a circuit can be significantly reduced, making it possible to design circuits and devices that would otherwise be impractical.

IC's are complete electronic circuits consisting of as many transistors, diodes, resistors, and capacitors as may be necessary for circuit operation. They are encapsulated in packages that are often no larger than a single transistor. The technology and materials used in the manufacture of IC's are basically the same as those used in the manufacture of transistors and other solid-state devices. In addition, IC's are manufactured for a wide variety of applications and, as a result, are used throughout the electronics industry. This trend will probably continue as manufacturers and engineers find more applications for this versatile family of devices.

## Advantages

The small size of the integrated circuit is its most apparent advantage. A typical IC can be constructed on a piece of semiconductor material that is less than one tenth of an inch square. Even when the IC is suitably packaged, it still occupies only a small amount of space. The first standard IC package was approximately one quarter of an inch long and one eighth of an inch wide. However, it later became apparent that in most applications this package was smaller than was actually necessary because the final size of the equipment was often dictated by other components which were much larger. Therefore, the IC's that are used today come in packages that are somewhat larger than the first ones that were developed.

Integrated circuits have been used extensively in the aerospace industry to reduce the size and weight of satellites, missiles, and other types of space vehicles. They have helped to reduce the size of complex computer systems and they are used in devices such as hearing aids and portable electronic calculators. In most applications where size and weight must be reduced to an absolute minimum, the IC can make a significant contribution.

The small size of the integrated circuit also produces other fringe benefits. The smaller circuits consume less power than conventional circuits and they cost less to operate. They generate less heat and therefore generally do not require elaborate cooling or ventilation systems. The smaller circuits are also capable of operating at higher speeds because it takes less time for signals to travel through them. This is an important consideration in the digital field where thousands of decision-making circuits are used to provide rapid solutions to various types of problems.

The integrated circuit is also more reliable than a conventional circuit that is formed with discrete components. This greater reliability results because every component within the IC is a solid-state device and these components are permanently connected together with thin layers of metal. They are not soldered together like the components in a conventional circuit and a circuit failure due to faulty connections is less likely to occur. Also, all of the components within the IC are simultaneously formed as opposed to conventional circuitry which is assembled in a step-by-step manner. Therefore, there is less chance of making mistakes during the assembly of an IC.

Integrated circuits are also thoroughly tested after they are assembled and only those devices which meet the required specifications are considered suitable for either military and space or industrial and commercial applications. This extensive testing of IC's, combined with the construction techniques previously described, produces a device that is highly reliable. In fact, when the IC is compared with conventional circuitry, it is often found to be thirty or even fifty times more reliable. This high degree of reliability has made the IC an important component for use in aerospace equipment and in complex digital computer systems. In each of these applications, a tremendous amount of circuitry is used and this circuitry must be reliable. Aerospace equipment, such as satellites and missiles, are sometimes required to operate for years without failure and a highly complex digital computer can be rendered inoperative if just one component fails to operate properly.

The use of integrated circuits can also result in a substantial cost savings. When IC's are manufactured in large quantities, they can often be sold at prices which are well below those of conventional circuits. Manufacturers of IC's usually offer a standard line of devices which are produced in large quantities. If designers can utilize those standard IC's in their equipment and if they purchase them in large quantities, they can often save a considerable amount of money. However, there are situations where the cost of an integrated circuit can be higher than that of a conventional circuit. This can occur when the designer needs a special purpose IC or one that is specially constructed by the IC manufacturer to suit the needs of the designer. Under these conditions, the cost might be prohibitive, especially when only a few IC's are purchased. However, as the IC manufacturers gain experience and manufacturing techniques improve, the cost of most types of IC's continue to decrease and this trend is likely to continue for some time to come.

By using integrated circuits in electronic equipment, additional cost savings can also be realized. Any equipment which utilizes IC's, instead of conventional circuitry, will have a fewer number of parts which must be assembled. Therefore, less wiring is required and less time is needed to assemble the equipment. This can mean a considerable cost reduction when a large number of units are to be produced. Also, the equipment manufacturer only has to procure and stock a relatively small number of IC's as compared to the relatively large inventory of discrete components that would be required if conventional circuitry was used. All of these factors can result in a reduction in overhead and cost savings for the equipment manufacturer, which can be passed on to the consumer.

## Disadvantages

It might appear that the integrated circuit has only advantages to offer and no real disadvantages. Unfortunately, this is not the case, since IC's do indeed have certain limitations which make them unsuitable for certain applications. Since the IC is an extremely small device, it cannot handle large currents or voltages. High currents generate heat within the device and the tiny components can be easily damaged if the heat becomes excessive. High voltages can break down the insulation between the components in the IC because the components are very close together. This can result in shorts between adjacent components which would make the IC completely useless. Therefore, most IC's are low power devices, which have low operating currents (in the milliampere range) and low operating voltages (5 to 20 volts). Also, most IC's have power dissipation ratings of less than 1 watt.

At the present time, only four types of components are commonly constructed within an IC, thus making available only a narrow selection of components. Diodes and transistors are the easiest to construct and are used extensively to perform as many functions as possible within each IC. Resistors and capacitors may also be formed, but it is much more difficult and expensive to construct these components. The amount of space occupied by a resistor increases with its value and in order to conserve space, it is necessary to use resistors with values that are as low as possible. It is also difficult to control the exact values of resistors, although the ratio between resistor values can be controlled with a high degree of accuracy. Capacitors occupy even more space than resistors and the amount of space required increases with the value of the capacitor. Therefore, capacitor values are made as small as possible and capacitors are used only when it is necessary. For these reasons, you will see diodes and transistors used extensively in IC's while resistors and capacitors are used sparingly.

Integrated circuits cannot be repaired because their internal components cannot be separated. When one internal component becomes defective, the whole IC becomes defective and must be replaced. This is definitely a disadvantage because it means that good components must be thrown away with the bad and it also means that there is the additional expense of replacing an entire circuit. However, this disadvantage is not as bad as it might first appear because it is offset by other factors which tend to compensate. First of all, the task of locating a trouble within a system is simplified because it is only necessary to trace the problem to a specific circuit instead of an individual component. This greatly simplifies the task of maintaining highly complex systems and reduces the demands that are placed on the maintenance personnel. Also, it is possible to reduce the time required to repair the equipment and the spare parts inventory can usually be smaller.

When all of the factors are considered, the disadvantages associated with the use of integrated circuits are outweighed by their advantages. There is a definite need for integrated circuits in the electronics industry. IC's are making it possible to reduce the size, weight, and cost of electronic equipment but at the same time increase its reliability. As manufacturing techniques improve, IC's are becoming more sophisticated and are capable of performing a wider range of functions. Therefore, in future years, these devices will undoubtedly find an increasingly broader range of military, space, industrial, and commercial applications.

# Self Test Review

1. An integrated circuit can be constructed on a piece of semiconductor material that is less than one tenth of an inch square.

   a. True
   b. False

2. Since the IC is extremely small it consumes only a small amount of ______________.

3. When IC's are used, the ______________ and ______________ of electronic equipment can be reduced.

4. Integrated circuits are capable of operating at higher ______________ than equivalent circuits using discrete components.

5. Since the IC is a single solid-state device, it is more ______________ than a conventional circuit which uses a variety of discrete components.

6. When IC's are manufactured in large quantities they usually cost ______________ than equivalent circuits using discrete components.

7. Integrated circuits cannot handle high operating ______________ or ______________.

8. The two IC components that are the easiest to construct are ______________ and ______________.

9. The two IC components that are the most difficult to construct are ______________ and ______________.

10. When a single component within an IC becomes defective, the entire IC must be replaced.

    a. True
    b. False

## Answers

1. (a) True
2. power
3. size and weight
4. speeds
5. reliable
6. less
7. currents or voltages
8. diodes and transistors
9. resistors and capacitors
10. (a) True

# IC CONSTRUCTION

As mentioned earlier, there are basically four types of integrated circuits. They are classified as monolithic, thin-film, thick-film, and hybrid devices. Each type has certain advantages over the others and also certain limitations. We will now see how each of these basic IC's are fabricated and we will also examine the most popular IC packages that are used.

## Monolithic IC's

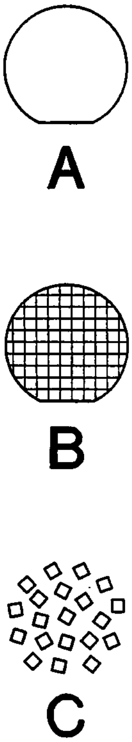
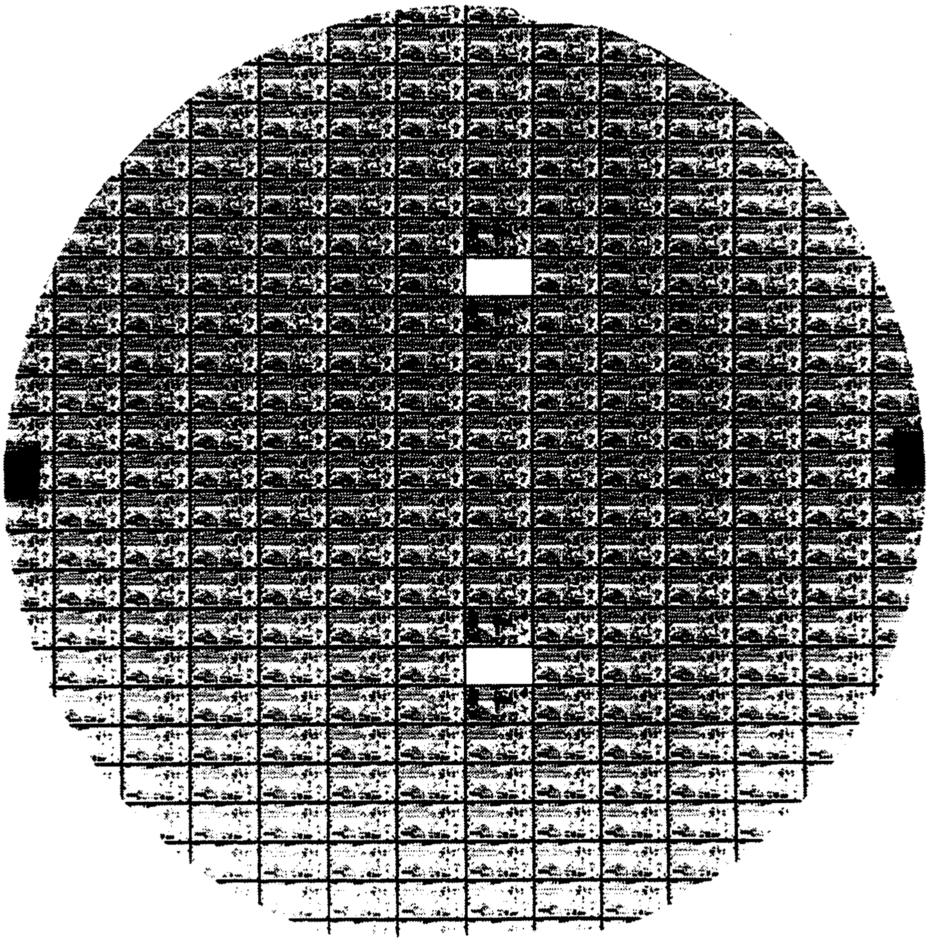


Figure 9-1
Integrated circuit construction.

The monolithic IC is constructed in basically the same manner as a bipolar transistor, although the overall process requires a few additional steps because of the greater complexity of the IC. The fabrication of a monolithic device begins with a circular semiconductor wafer (usually silicon) as shown in Figure 9-1A. This wafer is usually very thin (.006 to .012 inch thick) and may be as small as one inch or as large as two inches in diameter. The semiconductor wafer serves as a base on which the tiny integrated circuits are formed and is commonly referred to as a *substrate*.

Many IC's are simultaneously formed on a single wafer as shown in Figure 9-1B. Each square shown in this figure represents a single IC. The number of IC's formed on a wafer will depend on the size of the wafer and the size of the individual IC's. In some cases a wafer will contain less than 300 IC's but in other cases it may contain 500 to 600 IC's or possibly even more. All of the IC's formed on a wafer are usually the same type and therefore have the same physical dimensions. Also, each IC contains the same number and type of components.

When all of the IC's have been simultaneously formed, the wafer is sliced into many sections as shown in Figure 9-1C. These sections are commonly referred to as *chips* or *dice*. Each chip represents one complete integrated circuit and contains all of the components and wiring associated with that circuit. Once the IC's are separated into individual chips, each IC must be mounted in a suitable package and tested. Throughout this entire fabrication process certain IC's are rejected because they have various physical or electrical deficiencies. Although an extremely large number of IC's are fabricated in any given process, only a relatively small number of these devices are found to be usable. The number of usable devices is usually expressed as a percentage of the maximum number that is possible to obtain and is referred to as the *yield*. The yield in a typical fabrication process may be less than 20 percent. This simply means that less than 20 percent of the IC's produced in a particular process are found to be acceptable.



Over 200 monolithic IC's have been formed on this semiconductor wafer. The wafer is ready to be sliced into many individual IC's which will be subsequently packaged and tested. (Courtesy of Precision Monolithics)

Now that we have briefly reviewed the overall IC construction process, let's take a closer look at an individual monolithic IC and see exactly how the individual components within the IC are formed. As we examine the various IC components, you must remember that these components are formed within a number of IC's simultaneously on a semiconductor wafer. A single monolithic IC is never constructed alone.

## Bipolar IC's

As mentioned previously, the components that are commonly used in IC's are diodes, transistors, resistors, and capacitors. These components can be formed by diffusing impurities into selected regions of a semiconductor wafer (substrate) to produce PN junctions at specific locations. The basic manner in which these four components are formed and the manner in which they are interconnected are shown in Figure 9-2.

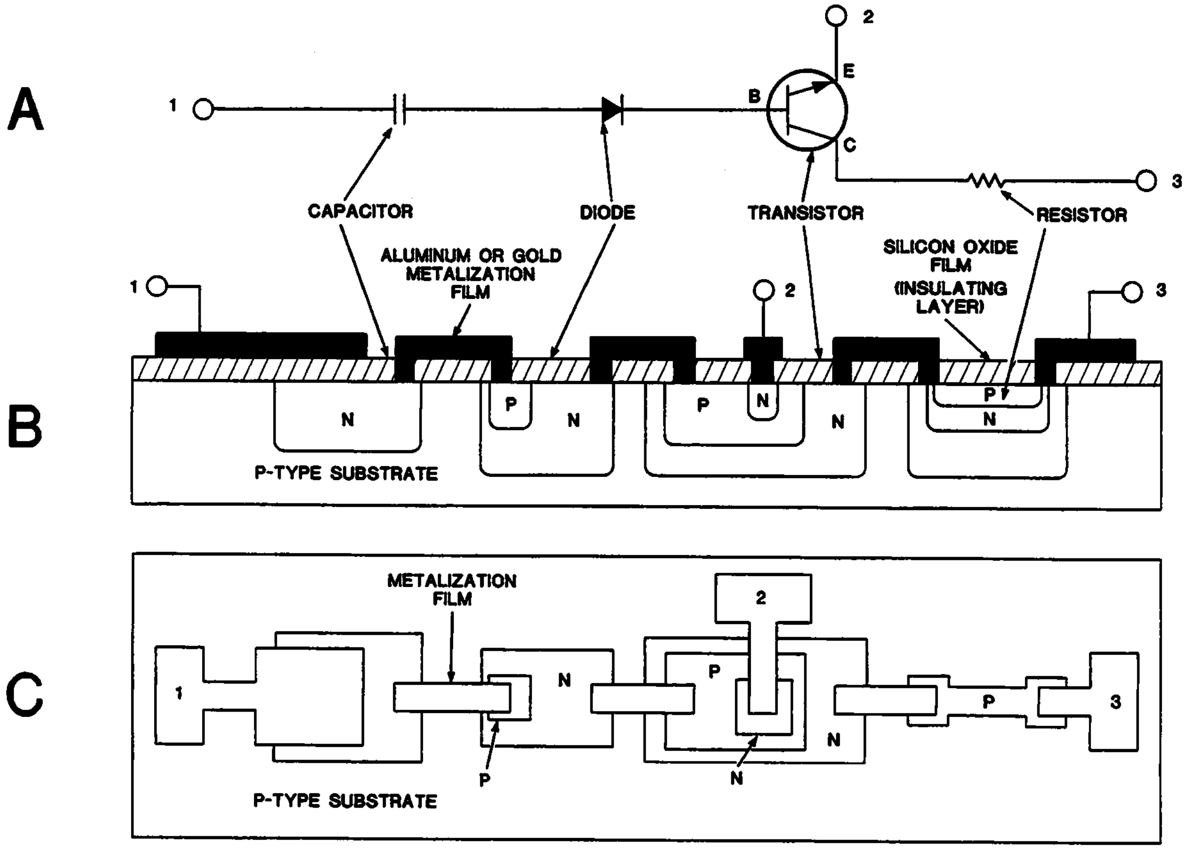


Figure 9-2
Basic construction
of a single integrated circuit.

## Definition

**Diffusion:** A popular construction technique which allows a gaseous N- or P-type impurity element, under high temperature, to penetrate a semiconductor wafer to produce a respective N- or P-type region.

Figure 9-2A shows a simple electronic circuit consisting of a capacitor, a PN junction diode, an NPN transistor and a resistor. Operating voltages and currents can be applied to the circuit through terminals 1, 2, and 3 as shown. This circuit could be easily constructed using four discrete components, however, it can also be produced as a monolithic integrated circuit. When produced in IC form, the circuit will appear as shown in Figure 9-2B and Figure 9-2C.

Observe the IC structure shown in Figure 9-2B and compare it with the circuit in Figure 9-2A. Notice that all four of the components are formed within a P-type substrate or wafer. These components are simultaneously formed by diffusing N-type and P-type impurities into the P-type substrate to produce N-type or P-type regions. Several diffusion operations are necessary to form the entire IC and it is necessary to accurately control the location, the size, and the depth of each N-type or P-type region that is formed. This is accomplished by depositing an insulating layer of silicon oxide on top of the substrate. Then appropriate windows are cut in the oxide coating by means of an acid so that only the desired areas on the substrate are exposed. Then N-type impurities are diffused through the windows and into the substrate to form the first and largest N-type regions. Next, the windows are covered with new oxide and new windows are formed. Then a P-type impurity is diffused into the substrate at the appropriate locations to form P-type regions. This process is repeated again to form the final N-type regions.

Notice that the NPN transistor in Figure 9-2B is constructed in basically the same manner as a conventional transistor. Its N-type collector was formed first. Then its P-type base and N-type emitter regions were formed. All three regions extend to the top of the substrate and lie in a flat plane, thus giving the device a *planar* type of structure. The transistor in Figure 9-2B is therefore equivalent to a conventional planar transistor that is formed with the diffusion method.

The PN junction diode in Figure 9-2 is also equivalent to a conventional diode. The larger N-type region, that was first diffused into the substrate, serves as the cathode and the smaller P-type region, that was diffused into the N-type region, serves as the anode.

The resistor is formed by first producing a large N-type region and then forming a P-type region on top of it. This produces a long narrow strip of P-type material that is surrounded by N-type material. This long P-type region serves as a resistor and its resistance can be controlled by adjusting its length and width. Increasing the length will result in a higher resistance, while increasing the width will lower the resistance. The resistance value can also be adjusted by controlling the concentration of impurities within the P-type strip. A higher concentration of impurities will result in a lower resistance and vice-versa.

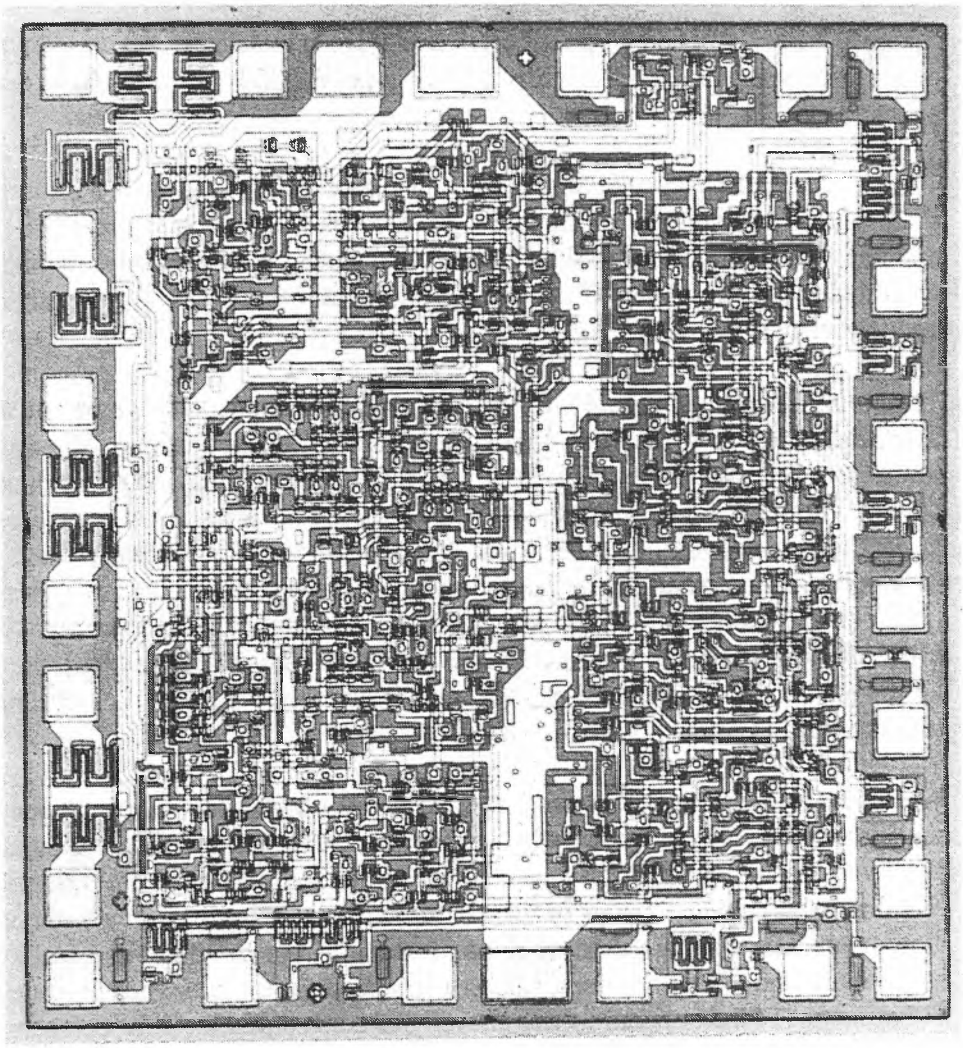
In general, the amount of space required to produce a resistor increases with its value, thus making it difficult to place large values within an IC. Also, it is difficult to produce resistors with highly accurate values. In general it is difficult to obtain resistance tolerances that are much better than plus or minus 10 percent, thus making it impossible to construct certain types or circuits in IC form. However, the ratio between two resistors that are formed close together on a substrate can be regulated with a high degree of accuracy. In fact, resistance ratios that are accurate to within plus or minus 1 percent can be obtained. Therefore, circuits that are produced in IC form are generally designed to take advantage of the accurate resistance ratios that are available, instead of specific resistance values.

The capacitor in Figure 9-2B is constructed in a unique manner. The diffused N-type region serves as the bottom plate of the capacitor and the layer of silicon oxide serves as the dielectric. The top plate of the capacitor is simply a layer of metal that has been deposited on top of the oxide coating. This type of capacitor is sometimes referred to as a *metal-oxide capacitor*.

The value of capacitance is determined by the area of the plates, the thickness of the oxide layer, and the dielectric constant of the oxide layer. In general, the amount of capacitance obtainable per unit area is very low and it is necessary to make the capacitor quite large to obtain a significant value of capacitance. In most cases, it is not practical to construct capacitors with values that are much higher than a few hundred picofarads. Even the smaller 10 and 20 picofarad capacitors require substantially more space than a transistor.

Another type of capacitor is also used in monolithic IC's. This device makes use of the capacitance that exists across the depletion region within a reverse-biased PN junction and it is commonly referred to as a *junction capacitor*. The junction capacitor is constructed in basically the same manner as the PN junction diode shown in Figure 9-2B. The amount of capacitance obtainable within the junction capacitor per unit area is also very low and, in general, the maximum capacitance values that can be obtained are somewhat lower than with metal-oxide devices. Also, the capacitance changes when the reverse bias voltage changes. This is because the bias voltage controls the size of the depletion region. This means that it is always necessary to maintain the proper bias voltage across the device.

As previously mentioned, all of the components in Figure 9-2B are simultaneously formed by coating the substrate with silicon oxide, cutting appropriate windows into the oxide coating, and diffusing impurities into the substrate. This process is repeated several times until all



A magnified view of a single monolithic integrated circuit. (Courtesy of Motorola Inc.)

components are formed. In the final steps of the process, an oxide film is deposited over the substrate and windows are cut into the oxide to expose the various regions within each component. Then a layer of aluminum or gold is deposited over the oxide and allowed to penetrate and make contact with the exposed regions. Next, acid is used to etch away certain portions of the metal film, leaving behind narrow strips of metal which serve as conductors and interconnect the various components to form a complete circuit. The end result of this operation can be seen in Figure 9-2B. A top view of the IC (with the silicon oxide removed) is also shown in Figure 9-2C. Notice that the four components are connected together by thin metal strips. The three terminals which provide access to the integrated circuit are also shown. Terminal 1 is simply the top plate of the capacitor which is formed simultaneously with the other metal strips. Terminals 2 and 3 are just metal strips which make direct contact with the transistor and resistor, respectively. As shown in Figure 9-2C, the metal strips which connect to terminals 1, 2, and 3 are brought out to the perimeter of the IC and enlarged to form rectangular pads. These pads (identified as 1, 2, and 3) provide relatively large metal surfaces to which fine metal wires can be bonded when the IC is permanently installed in a suitable package.

The components in the monolithic IC just described are formed by using the same basic construction techniques used to produce bipolar transistors. Integrated circuits that are produced in this manner are often called bipolar integrated circuits. Bipolar IC's are the most popular type of IC and are used in a wide variety of applications.

The monolithic IC in Figure 9-2 is extremely simple compared to most units that are constructed. Most IC's contain a large number of components and are therefore quite complex. In general, IC's are classified according to their complexity or the number of components that they contain. The terms *small-scale integration* (SSI), *medium-scale integration* (MSI), and *large-scale integration* (LSI) are commonly used to identify these circuits. Although these terms have never been exactly defined, SSI devices usually have less than 200 components while MSI devices generally have between 200 and 1000 components. When the IC has over 1000 components it is usually referred to as an LSI device. The simple IC shown in Figure 9-2 is an SSI device.

As you examine Figure 9-2, you may wonder why the various components do not interfere with each other when current flows through the circuit. Actually, the components are electrically isolated or insulated from each other due to the unidirectional characteristics of the PN junctions that are used. Due to the diffusion techniques that are used, each component is formed within its own N-type region and therefore each component is separated from the P-type substrate by a PN junction. When operating, the circuit is biased so that the P-type substrate is more negative than any other part of the circuit. This causes each PN junction to be reverse-biased and offer a high resistance which isolates the various components. The operating currents that flow through the circuit are therefore forced to take the proper paths. When IC's are constructed in this manner, they effectively utilize *diode isolation* as a means of electrically separating the various components.

Another construction technique is also sometimes used which makes use of *oxide-isolation*. With this method, a layer of silicon oxide is actually formed around each component, thus providing an insulating layer between each component and the substrate. A third method, known as *beam-lead isolation*, is also used in special applications. With this method, the components are formed as described earlier in the text, but a heavier metalization is used to form the interconnections and the semiconductor material between the components is removed. The components are therefore completely separated but supported by the heavy metalization that serves as the interconnecting conductors. A special type of plastic is then used to fill the spaces that were previously occupied by semiconductor material. With this method, an extremely high degree of isolation can be obtained.

## MOS IC's

To this point we have assumed that the transistors used in IC's are always bipolar devices. However, this is not the case. The fact is, IC's are often designed to utilize either bipolar transistors or field-effect transistors. The FET's that are used in IC's are the insulated-gate devices which are know as IGFET's or MOSFET's. These MOSFET's can be either N-channel or P-channel devices and they may operate in either the depletion or the enhancement mode.

The MOSFET's used in IC's are basically constructed like the conventional MOSFET's described in an earlier unit. A typical integrated circuit MOSFET is shown in Figure 9-3. Notice that the source and drain regions are diffused into the substrate. These regions will always be doped oppositely with respect to the substrate. A thin layer of silicon oxide is formed over the substrate and appropriate windows are cut into it so that metal electrodes (terminals) can be formed at the proper locations. Notice that the gate terminal is separated from the substrate by an extremely thin oxide layer. This oxide layer may be only 1000 angstroms thick (1 angstrom equals 1 × $10^{-10}$ meters), but it completely isolates the gate from the substrate. When properly biased, current will flow from source to drain through a channel that is formed directly under the gate, thus causing the device to operate in the enhancement mode.

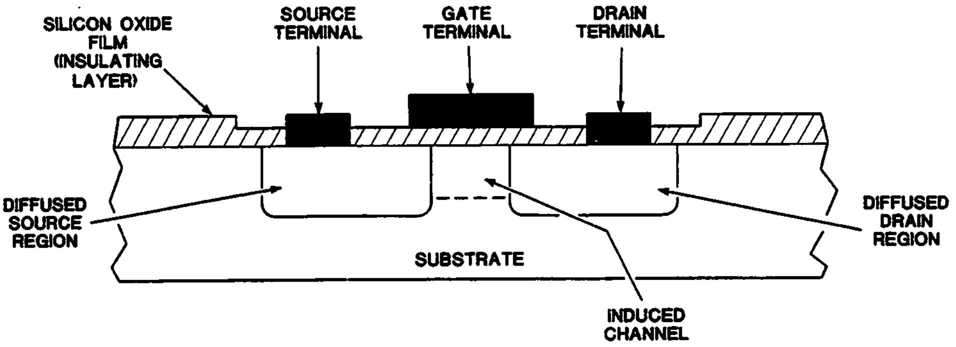


Figure 9-3
A typical integrated circuit MOSFET.

The source, drain, and gate regions within the MOSFET can be formed very close together. In fact, a typical MOSFET can be made much smaller than a bipolar transistor. This fact alone makes the MOSFET a very desirable component for use in integrated circuits. However, these devices have another distinct advantage. Due to the manner in which they are constructed, a high degree of isolation between components can be obtained. The gate is completely isolated by the oxide coating and the source and drain regions are isolated by PN junctions.

The MOSFET can also be used as a resistor when it is properly biased. The resistance of the channel that is formed between the its source and drain terminals can be adjusted by regulating its operating voltages. This resistance is also determined by the transconductance of the device. When MOSFET's are used as resistors, a wide spacing is used between the source and drain regions to obtain a lower transconductance and the gate is connected directly to the drain so that the device is always conducting.

Integrated circuits which use MOSFET's as the principle controlling elements are generally referred to as MOS IC's. Many MOS IC's contain only MOSFET's while others contain combinations of MOSFET's and MOSFET resistors. MOS IC's are produced with the same basic technology that is used to produce field effect transistors, while bipolar IC's are based on bipolar transistor technology.

Since MOS components are smaller than bipolar components, the MOS circuit can be constructed in a smaller space than an equivalent bipolar circuit. This higher component density in MOS circuits makes them highly suitable for use in MSI and LSI circuits. MOS circuits also consume less power, provide greater temperature stability, and offer a higher input impedance than comparable bipolar circuits. Unfortunately, they have one important disadvantage which limits their use. In general, MOS circuits cannot respond as quickly as bipolar circuits. When used in digital or logic applications or in any other applications where basic switching actions are required, they cannot operate as fast as bipolar circuits. Even when used to amplify or generate electronic signals, MOS circuits cannot operate at the high frequencies which are obtainable with bipolar circuits. Therefore, in certain applications bipolar IC's may be more desirable than MOS circuits while in other cases the exact opposite will be true.

# Thin-Film IC's

Unlike monolithic IC's, which are formed within a semiconductor material (substrate), the thin-film circuit is formed on the surface of an insulating substrate. In the thin-film circuit, components such as resistors and capacitors are formed from extremely thin layers of metals and oxides which are deposited on a glass or ceramic substrate. Interconnecting wires are also deposited on the substrate as thin strips of metal. Components such as diodes and transistors are formed as separate semiconductor devices and then permanently attached to the substrate at the appropriate locations.

The substrate on which the thin-film circuit is formed is usually less than one inch square. The resistors are formed by depositing tantalum or nichrome as thin films or strips on the surface of the substrate. These films are usually less than .0001 inch thick. The value of each resistor is determined by the length, width and thickness of each strip that is formed on the substrate. Several thin-film resistors and their interconnecting conductors are shown in Figure 9-4. The short, wide resistor has a relatively low value while the longer and narrower resistors have higher values. The interconnecting conductors are extremely thin metal strips which have been deposited on the substrate. Low resistance metals, such as gold, platinum, or aluminum, are generally used as conductors. The substrate is made from an insulating material that will provide a rigid (nonflexible) support for the components. Glass or ceramic materials are often used as substrates.

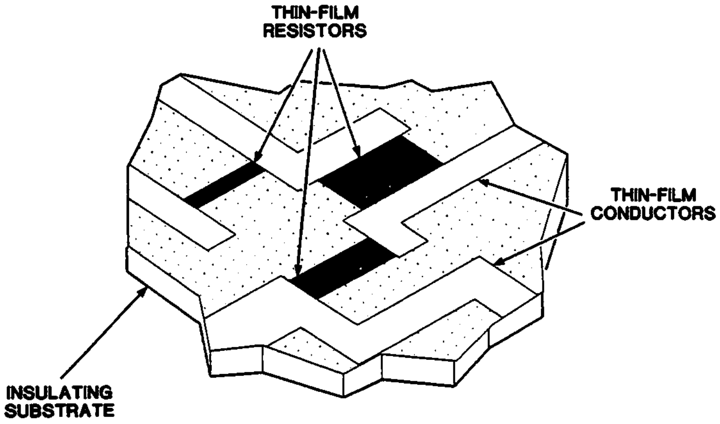


Figure 9-4
A portion of a thin-film circuit
showing resistors and conductors.

With thin-film techniques, it is possible to produce extremely accurate resistance values. The tolerances on resistor values can be as low as ± 0.1 percent over a resistance range that extends from several ohms to more than 100 K ohms. Furthermore, it is possible to obtain extremely accurate ratios between various resistors in a thin-film circuit. Ratio accuracies of ± 0.01 percent can be readily obtained.

Thin-film capacitors consist of two thin layers of metal separated by an extremely thin dielectric layer. One metal layer is deposited on the substrate and then an oxide coating is formed over the metal to serve as a dielectric. Next, the top plate of the capacitor is formed by depositing a thin metal film over the dielectric. Materials such as tantalum, gold or platinum can be used to form the plates and the dielectric may be formed from insulative materials such as tantalum oxide, silicon oxide, or aluminum oxide. The required capacitance value is obtained by adjusting the area of the plates and by varying the thickness as well as the type of dielectric material used.

When diodes or transistors are required in a thin-film circuit, they are produced as separate semiconductor components using the same basic monolithic techniques that are used to form conventional transistors. Many of the diodes and transistors that are used in thin-film circuits are formed by diffusion methods and have a planar type of construction. The diode and transistor chips are permanently mounted on the thin-film substrate and then electrically connected to the thin-film circuit with extremely thin wires.

The materials used in the construction of thin-film components and conductors are usually deposited on an insulating substrate by using either an *evaporation* or *sputtering* process. In the evaporation process, the materials are placed within a vacuum and heated until they evaporate. The resulting vapor is allowed to condense on the substrate to form a thin film. The sputtering process takes place in a gas filled chamber. High voltages are used to ionize the gas and the material is bombarded with ions. The atoms within the material become dislodged and drift toward the substrate where they are deposited in the form of a thin film. To insure that the proper films are deposited at the proper locations on the substrate, various masks can be used which will expose only the desired regions, or the substrate can be completely coated with a film and then the undesired portions can be cut or etched away.

# Thick-Film IC's

Thick-film IC's are formed in a somewhat different manner than the thin-film devices just described. In thick-film IC's, the resistors, capacitors and conductors are formed on an insulating substrate by using a *silk-screen* process. In the silk-screen process, a very fine wire screen is placed over the substrate and a metalized-ink is forced through the screen with a squeegee. Only certain portions of the wire screen are open, thus allowing the ink to penetrate and coat specific portions of the substrate. The remaining holes in the screen are filled with a special emulsion. In this manner, a pattern of interconnecting conductors is formed on the substrate. The pattern is then heated to over 600 degrees centigrade and the painted surfaces harden and become low resistance conductors. Resistors and capacitors are also silk-screened on top of the substrate by forcing the appropriate materials (in paste form) through an appropriate screen and then heating the substrate to a high temperature. This process is repeated using various pastes until the circuit is complete except for diodes or transistors. As with thin-film circuits, these components must be formed separately as semiconductor devices and then added to the substrate to complete the circuit.

Subsequent operations may also be required after the silk-screening is completed. For example, it is generally necessary to trim the resistors to obtain accurate resistance values. This trimming operation can be performed with an air-abrasive (sand blasting) technique or a laser can be used to burn away excessive resistor material. This trimming operation can produce resistors with tolerances as low as ± 0.5 percent and these resistors can be produced in standard values that can extend from 5 ohms to 100 megohms. Thick-film capacitors have relatively low values which can typically range up to several thousand picofarads. When higher capacitance values are required, the film type capacitors cannot be used. To obtain higher capacitance values, it is necessary to use miniaturized discrete capacitors and mount them permanently on the substrate.

The thick films formed in the silk-screening process are usually more than 0.0001 inch thick. Thick-film components are therefore larger than the thin-film components described earlier and come closer to resembling conventional discrete components. The tolerances involved in the construction of thick-film circuits are therefore not as critical as those associated with thin-film circuits. Thin-film components actually have no discernible thickness and in many cases appear to be simply painted or printed on a substrate.

## Hybrid IC's

Hybrid integrated circuits are formed by utilizing various combinations of monolithic, thin-film and thick-film techniques and they may even contain discrete semiconductor components in chip form. Therefore, many types of hybrid circuit arrangements can be produced. A typical hybrid circuit might consist of a thin-film circuit on which various monolithic IC's have been attached or it could utilize monolithic IC's, thick-film components and discrete diodes and transistors that are all mounted on a single insulating substrate.

A portion of a typical hybrid IC is illustrated in Figure 9-5. An insulating substrate is used to support the circuit components as shown. Notice that a monolithic IC is mounted on the substrate along with thick-film resistors and a small discrete capacitor. All of the components are interconnected with conductors that are formed on the substrate using film techniques. The monolithic IC is connected to the conductors with fine wires that are bonded in place. The thick-film resistors will usually have notches cut into them, as shown, as a result of the trimming process that is used to adjust their values. The capacitors used in these circuits can be formed either by using film techniques or miniature devices can be installed between conductors as shown. The conductors are routed to the

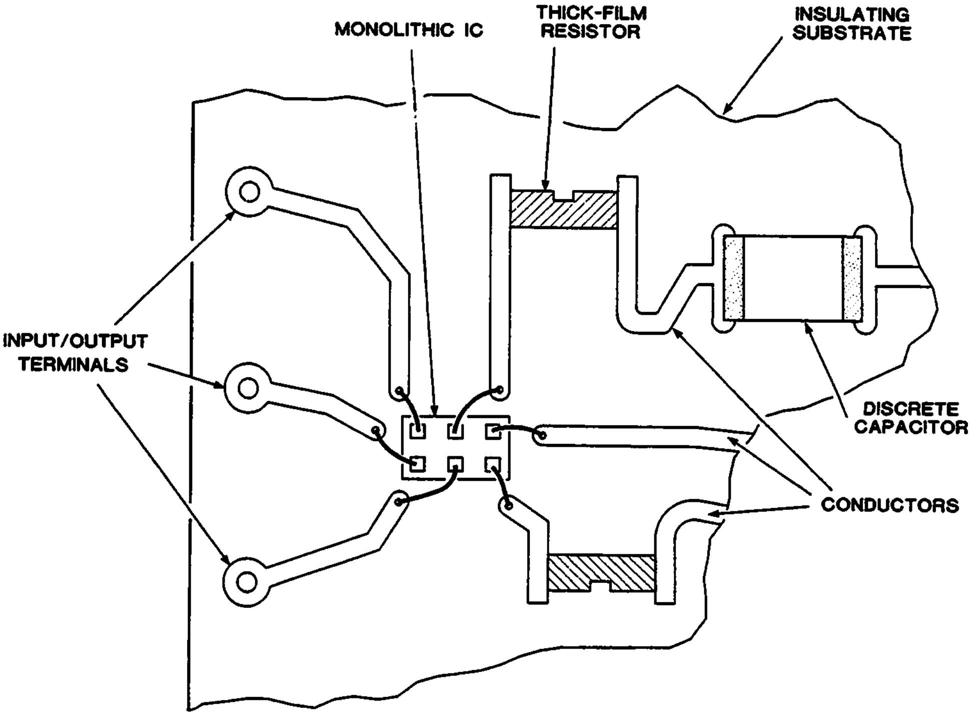


Figure 9-5
A portion of a typical hybrid IC.

outer edge of the substrate, as shown, where they are enlarged to form circular access terminals for the circuit. When the hybrid circuit is packaged, suitable pins are soldered to these terminals so that the circuit can be plugged into a matching receptacle or socket.

The hybrid approach to the construction of IC's allows a high degree of circuit complexity (a large number of components) by using monolithic circuits and at the same time takes advantage of the extremely accurate component values and tolerances that can be obtained with film techniques. Discrete components such as diodes, transistors, or capacitors are often used in hybrid circuits because they can handle relatively large amounts of power and because they exhibit certain desirable electrical characteristics that are difficult to achieve in a single monolithic circuit.

If only a few circuits are required, it can be much cheaper to utilize hybrid IC's instead of monolithic IC's. This is because the construction of hybrid circuits does not involve the elaborate and expensive circuit layout design procedures and diffusion techniques associated with monolithic circuits. The major expense in hybrid circuit construction is in the wiring and assembly of the components and in the final packaging of the devices. Therefore, hybrid circuits can be more easily designed and packaged for special applications. However, when a large number of circuits are required, monolithic devices are usually the better choice, if they can perform the required functions. Monolithic circuits are considerably less expensive when they are manufactured and sold in large quantities so that the initial design expense is overcome.

Since hybrid circuits utilize discrete components as well as monolithic and film circuits, they are often larger and heavier than monolithic IC's. The use of discrete components also tends to make them somewhat less reliable than monolithic circuits.

# IC Packages

Like transistors and other types of solid-state components, IC's are mounted in packages which protect them from moisture, dust and other types of contaminants. The packages also make it easier to install the IC's in various types of equipment, since each package contains leads which can be either plugged into matching sockets or soldered to adjacent components or conductors.

Many different types of IC packages are available and each type has its own advantages and disadvantages. The most popular IC package is the *dual in-line package*, which is commonly referred to as a DIP. A typical dual in-line package is shown in Figure 9-6. Notice that the package has two rows of mounting pins or leads which can be inserted in a matching socket, or inserted in holes on a printed circuit board and soldered in place. The package shown has 12 leads in each row or a total of 24 pins. However, both larger and smaller dual in-line packages are available to accommodate the wide range of SSI, MSI and LSI circuitry.

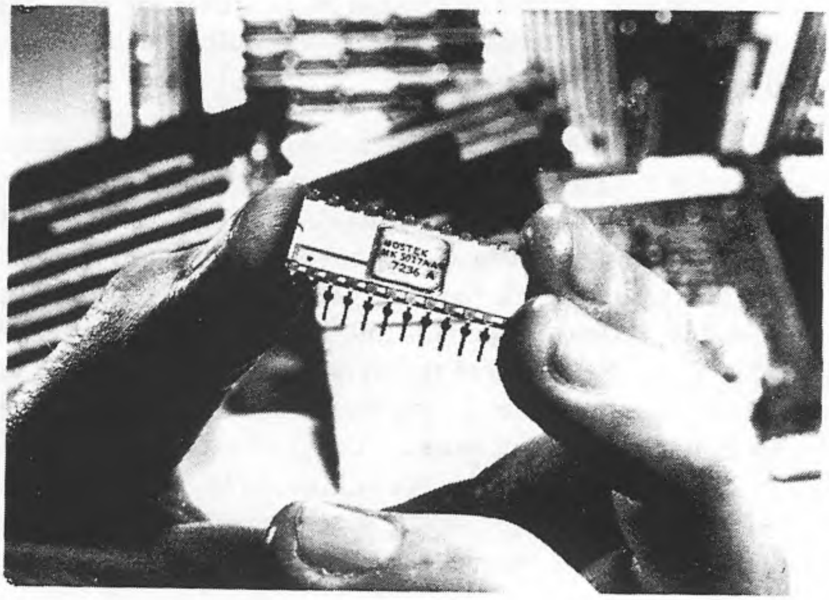


Figure 9-6
A typical dual in-line integrated circuit package
(Courtesy of MOSTEK.)

Dual in-line packages may be constructed from either plastic or ceramic materials. Plastic devices are relatively inexpensive and are considered suitable for most commercial and industrial applications where operating temperatures fall within a range of 0° to 70° centigrade. The ceramic devices are somewhat more expensive but offer better protection against moisture and contaminants. They can also withstand a wider range of operating temperatures (−55° to +125° centigrade) and are generally recommended for use in military, aerospace and severe industrial applications. Some of the most popular DIP's are made of plastic and have a total of 14 or 16 leads and some smaller versions with only 8 leads are also used. These smaller devices are commonly referred to as miniature dual in-line packages or mini DIP's. Several mini DIP's are shown in Figure 9-7.

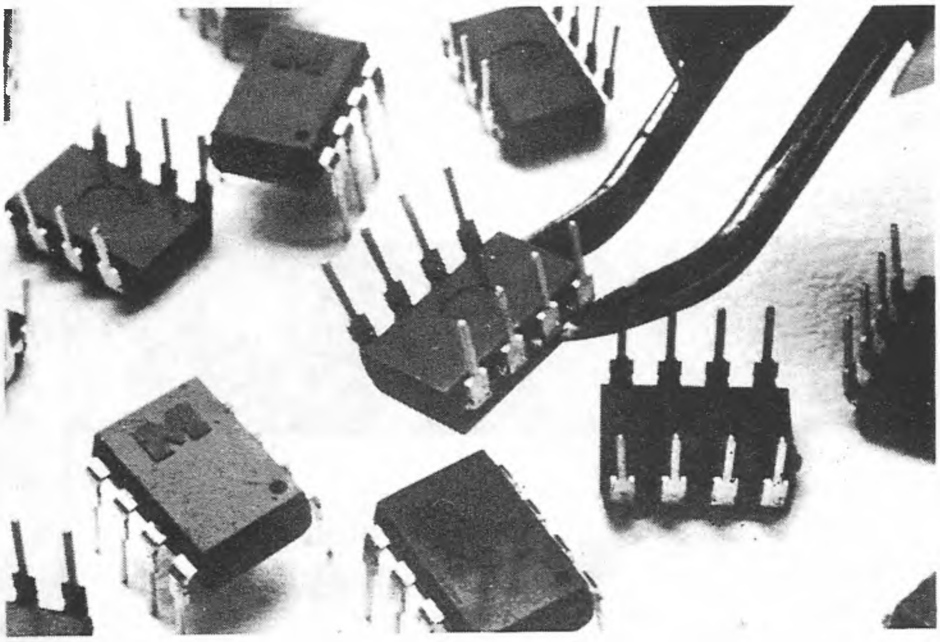


Figure 9-7
Typical miniature dual in-line
packages (Courtesy of Monsanto).

Dual in-line packages are most commonly used with monolithic integrated circuits, but they can also be used with other types of IC's. The tiny IC chips are permanently mounted inside of their respective packages and electrical connections between the chips and package leads are made with fine wires that are bonded in place. Although it is common practice to use one IC chip in a package, some complex circuits must be formed by mounting several chips in a single package and interconnecting them.

The large dual in-line package shown in Figure 9-8 contains three monolithic chips. A network of conductors have been formed on the same base that supports the chips. The various conductor pads on the chips are connected to these conductors with fine gold wires that have been bonded in place. The conductors, in turn, are connected to the two rows of leads along the edge of the package. The lid (or cover) shown immediately below the package, is placed over the opening in the package and soldered into place to provide an air tight (hermetically sealed) unit.

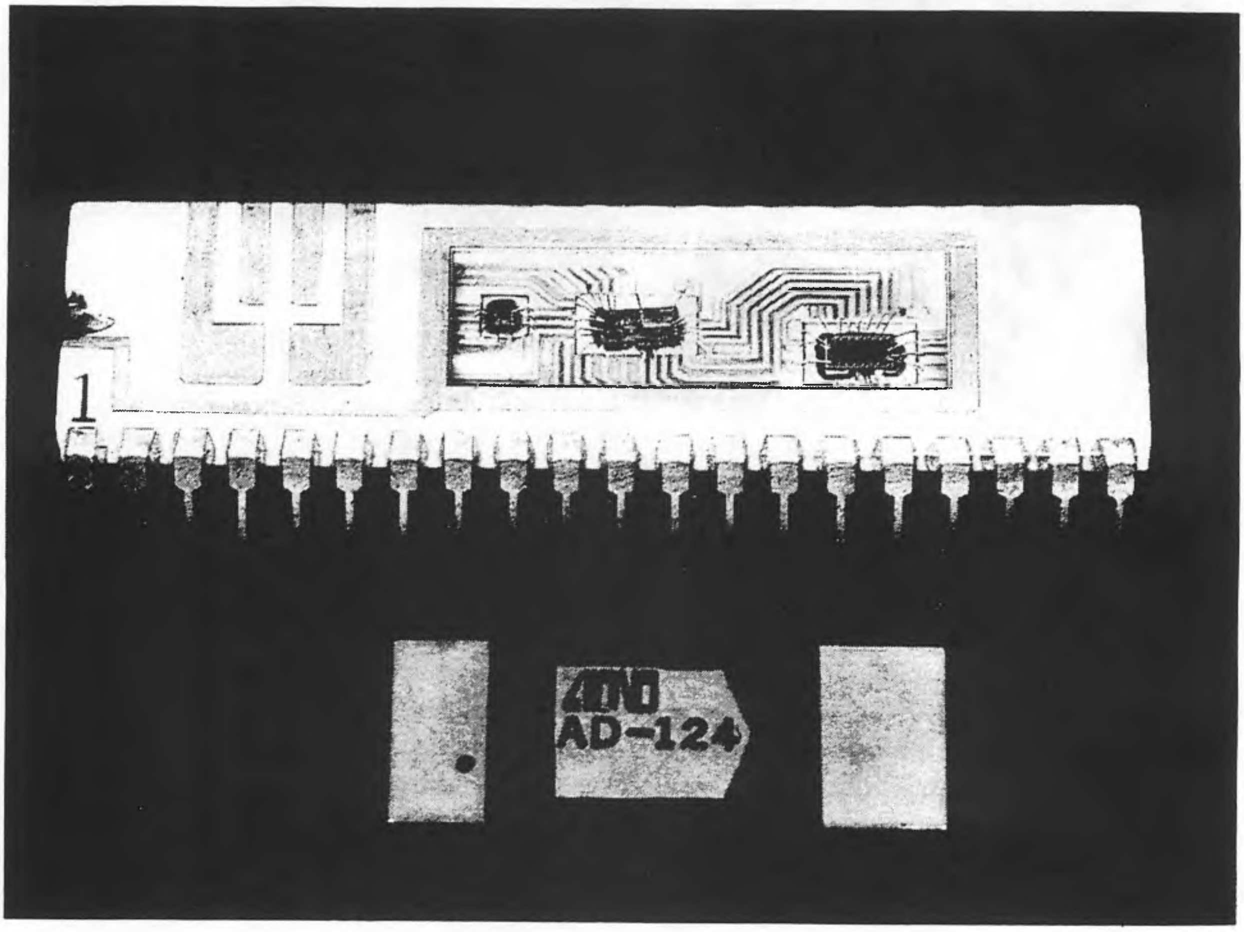


Figure 9-8
A large dual in-line package which contains three monolithic chips (Courtesy of Precision Monolicthics).

Another type of IC package that is widely used, is the *flat-pack.* The flat-pack is similar to the dual in-line package but it is smaller and much thinner. A typical flat-pack is illustrated in Figure 9-9A. Notice that the device is very thin and its leads extend horizontally outward around its edges. The flat-pack can be mounted almost flush with the surface of a printed board and its leads are usually soldered directly to adjacent conductor pads on the circuit board. The flat-pack is therefore used where space is limited. It is often made from metal or ceramic materials and can be used over a wide range of operating temperatures (-55° to + 125° centigrade). The popular units often have a total of 10 or 14 leads, although larger and smaller units are also available. Flat-packs are widely used with monolithic circuits to take full advantage of their extremely small size.

Integrated circuits may also be mounted in *metal cans* that are similar to the types used to house transistors. A typical metal can is shown in Figure 9-9B. The metal can shown has only 8 leads, although larger devices with more leads are available. Metal cans may be used to house monolithic film-type or hybrid IC's and their long metal leads make it possible to install them in a variety of ways. Metal cans may be used over a wide temperature range (–55° to +125° centigrade) and are therefore suitable for military and space applications.

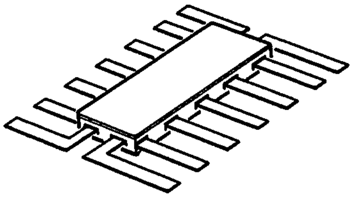


Figure 9-9A

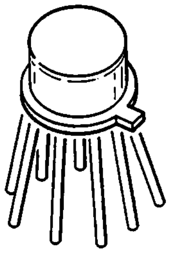


Figure 9-9B

Hybrid integrated circuits and film circuits are constructed on insulating substrates (usually ceramic materials), which in some cases serve as part of their final package. For example, the substrate shown in Figure 9-5 has input and output terminals around its outer edge to which pins can be attached. With the pins attached, the substrate can be plugged into a matching socket or plugged into a printed circuit board and soldered into place. All that is needed to complete the package is a ceramic lid or cover that can be placed on top of the substrate to protect the components.

After the integrated circuits are packaged in their respective containers, they are put through a series of tests to be sure that they meet certain electrical specifications. Since the performance of any circuit is affected by changes in temperature, the IC's must be tested over a wide range of operating temperatures to be sure that they are suitable for either commercial and industrial applications, or military and space operations.

Surface mount technology (SMT) has brought about the development of four (4) new integrated circuit packages. The small outline integrated circuit (SOIC), the plastic leaded chip carrier (PLCC) shown in Figure 9-10, the ceramic leaded chip carrier (LDCC) and the leadless chip carrier (LCC).

IC's having 4-16 terminals are usually supplied in the SOIC package. The PLCC is often used for IC's having more than 18 terminals. The LDCC is similar to the PLCC, but operates at higher temperatures. The LCC's have no leads. Instead of leads, leadless terminals border the sides of the device.

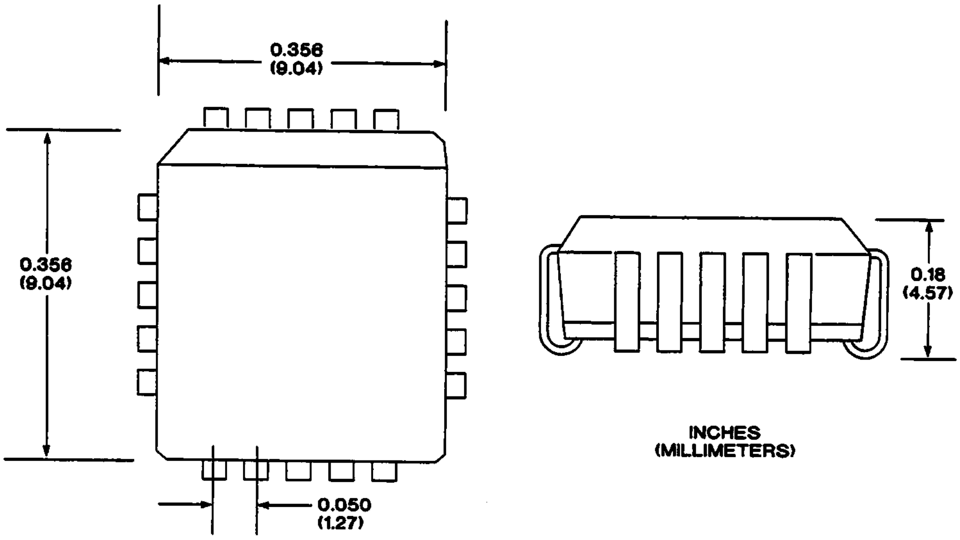


Figure 9-10
Outline of a PLCC-20 square 20-pin PLCC.

## Self Test Review

11. The monolithic IC is produced by using the same basic techniques that are used to produce a bipolar transistor.

    a. True
    b. False

12. As many as 500 IC's can be simultaneously formed on a circular semiconductor _______________.

13. The semiconductor material on which the IC's are formed is commonly referred to as a _______________.

14. When the monolithic IC's have been formed they are sliced in individual sections which are called _______________ or _______________.

15. The number of usable IC's obtained in a given process is usually expressed as a percentage of the maximum number possible and is referred to as the _______________.

16. The value of a monolithic IC resistor can be adjusted by physically varying its _______________ and _______________.

17. The two types of capacitors often used in monolithic IC's are referred to as _______________ and _______________ capacitors.

18. In general, the size of a monolithic resistor or capacitor increases as its value _______________.

19. Integrated circuits are classified according to complexity or the number of components they contain and are identified as _______________, _______________, or _______________ circuits.

20. The monolithic IC in Figure 9-2 utilizes _______________ isolation as a means of electricity separating components.

21. Monolithic components may also be separated by means of ____________ isolation or _______________ isolation.

22. Monolithic IC's that utilize bipolar transistor construction techniques as shown in Figure 9-2, are known as _______________ IC's.

23. Monolithic IC's that utilize MOSFET's as the principle controlling elements are referred to as _______________ IC's.

24. Thin-film IC's are formed by depositing metals and oxides in extremely thin layers on an ______________ substrate.

25. With thin-film techniques, it is possible to produce extremely accurate resistance values.

    a. True
    b. False

26. Thin-film techniques can be used to produce diodes and transistors, as well as resistors and capacitors.

    a. True
    b. False

27. The materials used to form thin-film circuits are deposited by using an ______________ or ______________ process.

28. In thick-film circuits, resistors, capacitors and conductors are formed on an insulating substrate with a ____________ process.

29. Hybrid IC's are devices which contain combinations of ______________, ______________ and ______________ circuits.

30. Three integrated circuit packages that are widely used are the __________ package, the ________________, and the ________________.

## Answers

11. (a) True
12. wafer
13. substrate
14. chips or dice
15. yield
16. length and width
17. metal-oxide and junction
18. increases
19. SSI, MSI or LSI
20. diode
21. oxide isolation or beam-lead isolation
22. bipolar
23. MOS
24. insulating
25. (a) True
26. (b) False — Thin-film techniques are used to produce resistors and capacitors. Diodes and transistors must be produced separately with semiconductor materials and then attached to thin-film circuits.
27. evaporation or sputtering
28. silk-screen
29. monolithic, thin-film and thick-film
30. dual in-line package, the flat-pack, and the metal can.

# APPLICATIONS OF IC's

Integrated circuits are placed into two general categories, digital IC's, or linear IC's. Digital IC's are the most widely used devices. They are designed for use in various types of logic circuits, such as in digital computers. A linear IC provides an output signal that is proportional to the input signal applied to the device. Linear IC's are widely used to provide such functions as amplification and regulation. They are used in television sets, FM receivers, electronic power supplies, and in various types of communications equipment.

It is impossible to consider all of the possible applications of digital and linear IC's. However, we will examine a few typical examples. First, we will examine digital IC's.

## Digital IC's

Digital circuits use discrete values (either 0 or 1) to perform 3 general functions. They are the AND, OR, and NOT functions. The 3 functions are performed by logic circuits that are called the AND, OR, and NOT logic gates. These gates or circuit configurations can be combined to make decisions based on digital input information. Thus, the gates are also called digital logic gates.

We will define each of the 3 basic functions. This will be accomplished using truth tables that show all of the possible inputs and outputs of the AND, OR and NOT logic gates. In a digital gate it is only possible to have an output of either 0 or 1. Zero and 1 are representations for voltage levels. Usually a voltage of approximately 3 volts or higher is considered to be a logic 1 and voltages of 0.7 and below are considered to be logic 0.

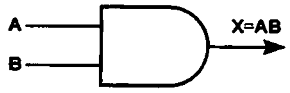
## AND Gate

Figure 9-11 shows the AND truth table and logic symbol. The AND truth table shows the possible input (A and B) and output (x) combinations.

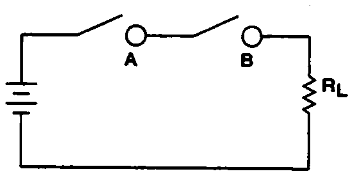
The AND gate has an output of 1 only when all of its inputs are equal to 1. This is similar to a multiplier function since the only possibilities in a digital circuit are 0 × 1 = 0 and 1 × 1 = 1.

| A | B | X |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

AND TRUTH TABLE



AND LOGIC SYMBOL



AND SCHEMATIC DIAGRAM

Figure 9-11
AND gate truth table, logic symbol, and schematic diagram.

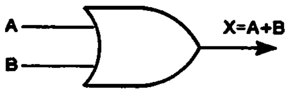
The schematic drawing in Figure 9-11 has 2 switches connected in series. Unless both switches are closed, there is no current flow to the output.

## OR Gate

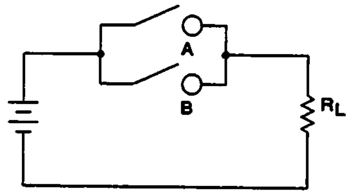
Figure 9-12 shows the OR truth table and logic symbol. The truth table indicates possible inputs (A and B) and output (x) combinations. The OR gate provides a 1 in the output when any or all of its inputs are 1. This is similar to an adder function except that it doesn't provide for a carry. Thus, 0 + 0 = 0, 0 + 1 = 1, and 1 + 1 = 1. The + symbol doesn't really mean add, it means that A and B are ORed together.

| A | B | X |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

OR TRUTH TABLE



OR LOGIC SYMBOL



OR SCHEMATIC DIAGRAM

Figure 9-12
OR gate truth table, logic symbol, and schematic diagram.

The schematic drawing in Figure 9-12 has 2 switches connected in parallel and if either or both switches are closed there is a complete path for current flow.

## NOT Gate

The NOT gate provides an output that is always opposite the input. This is called inversion or 180 degrees phase shift. Thus, the NOT gate is commonly referred to as an **inverter**. Remember that the common emitter amplifier configuration, Figure 9-13, was the only amplifier configuration capable of inverting the input signal. Therefore, the common emitter configuration is capable of performing the NOT function and when used as an inverter it is driven between saturation and cutoff. Figure 9-13 also show the logic symbol and truth table.

| A | $\bar{A}$ |
|---|---|
| 0 | 1 |
| 1 | 0 |

TRUTH TABLE

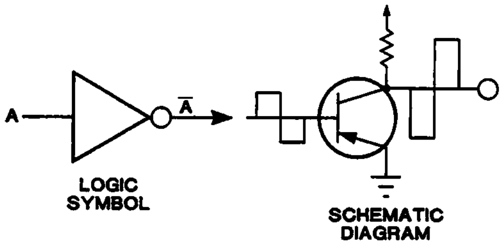
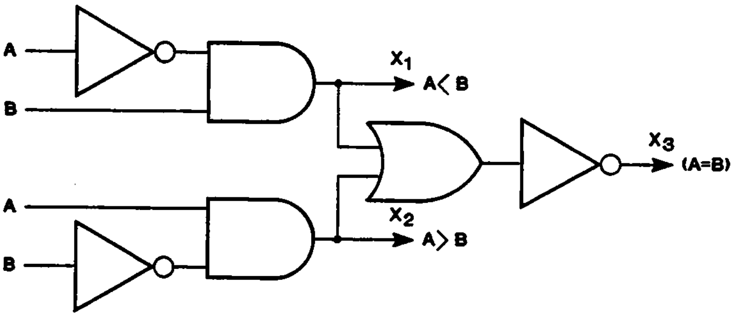


Figure 9-13
NOT gate truth table, logic symbol, and schematic diagram.

## Combinational Logic Circuit

This circuit combines the AND, OR, and NOT functions into a single decision making circuit with 3 distinct outputs.



COMBINATION LOGIC CIRCUIT

| A | B | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 1 |

TRUTH TABLE

Figure 9-14
Combined logic symbols.

In Figure 9-14 the circuit uses 2 AND gates, 1 OR gate, and 3 NOT gates to provide outputs when A is less than B, A is greater than B, and when A and B are equal. This is sometimes referred to as a choice box in a digital computer program.

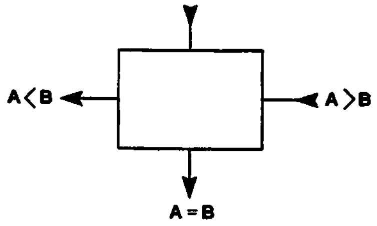


Figure 9-15
Logic symbol for decision making circuit.

## Typical ICs

Most of the integrated circuits in use today are digital IC's. These devices are widely used in digital computers and portable electronic calculators to perform various arithmetic and decision making functions. Digital IC's are produced using both bipolar and MOS construction techniques. These circuits can be very simple or extremely complex and are therefore available at the SSI, MSI, and LSI levels.

A typical digital IC which is formed by using bipolar construction techniques is shown in Figure 9-16. A schematic diagram of the IC is shown in Figure 9-16A. Notice that only transistors, diodes, and resistors are used in the circuit and since it contains only 11 components, it is classified as an SSI circuit. In this circuit, the transistors are the key elements and because of the unique manner in which they are connected, the circuit is commonly referred to as a transistor-transistor logic (TTL) circuit.

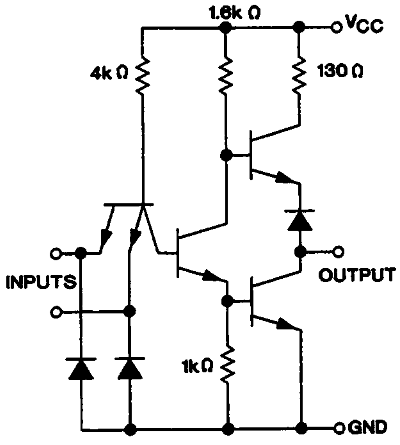
The TTL circuit in Figure 9-16A performs an important logic function. It is capable of comparing two input voltage levels, which must be equal to either 0 volts or approximately 3 volts, and provide an output voltage level (0 or 3 volts) depending on the input combination. The circuit performs what is commonly referred to as the NAND function and the circuit itself is referred to as a NAND gate since it provides a gating or switching action between two voltage levels. The NAND gate, like all digital circuits, is capable of recognizing only two voltage levels (sometimes called logic levels) at each of its inputs. Instead of referring to the specific voltages involved (which can vary with different types of digital circuits), it is common practice to refer to one level as a high logic level or a logic 1, and the other voltage level as a low logic level or a logic 0. A digital circuit can therefore be thought of as a device which responds to various high and low (1 or 0) logic levels and the actual voltages involved can be ignored. The table in Figure 9-16B shows the output levels (1 or 0) produced by the NAND gate when all possible combinations (only 4 are possible) of input levels have been applied. The NAND gate is therefore capable of making a simple decision based on the combinations of logic levels at its inputs and provide a specific output logic level for each combination.

Since many thousands of NAND gates are used in digital computers and other complex digital systems, it is not practical to draw the entire circuit each time it is shown on a schematic. Therefore, the NAND gate is usually represented by the symbol shown in Figure 9-16C. Notice that only the two inputs and the output of the circuit are represented.

It is common practice to construct not one but four of these NAND gate circuits on a single IC chip, using bipolar techniques, and mount the chip in a single package. Both dual in-line packages and flat-packs are widely used with IC's of this type. The package outline of a typical dual in-line package is shown in Figure 9-16D. This package outline drawing shows how the various NAND gates are internally connected to the package leads. Notice that the IC package has 14 lead (also called pins), which are consecutively numbered in a counterclockwise direction. The package also has a notch at one end, which serves as a key to help locate pin number 1. Notice that pins 1 and 2 serve as inputs to one gate and pin 3 provides the output connection. Power is simultaneously applied to all four circuits through pins 14 and 7.

The package outline drawing shown in Figure 9-16D is typical of those provided by IC manufacturers in their specification sheets. They may also provide a schematic of the particular circuit involved and, of course, will always provide the important electrical characteristics of the circuit. In many cases, the circuit designer or engineer is more interested in what the circuit can do and is less interested in how it does it or how it is constructed. Therefore, specification sheets are likely to contain more mechanical and electrical information which pertains to the overall performance of the IC and very little information relating to its internal construction.

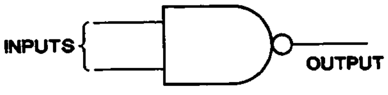
A typical example of a digital IC which is formed by using MOS construction techniques is shown in Figure 9-17. This IC utilizes one of the newest and most advanced construction techniques. It contains both P-channel and N-channel enhancement mode MOS field-effect transistors and is commonly referred to as a complementary-symmetry/metal-oxide semiconductor IC or simply a CMOS IC. CMOS circuits containing P- and N-channel MOSFET's are now widely used because they have many advantages over other types of digital circuits. They consume less power than other types of digital IC's and they have good temperature stability. They can operate over a wide range of supply voltages (typically 3 to 15 volts), as compared to TTL circuits which require an accurate 5 volt supply. CMOS circuits also have a high input resistance, which makes it possible to connect a large number of circuit inputs to a single output without loading down the output and disrupting circuit operation. This is an extremely important advantage in digital equipment where thousands of circuit are used.



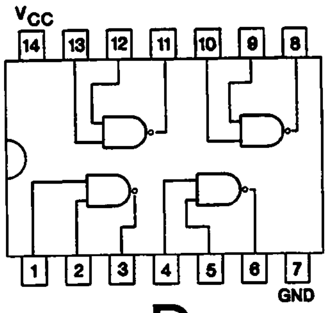
A

| INPUTS | | OUTPUT |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

B



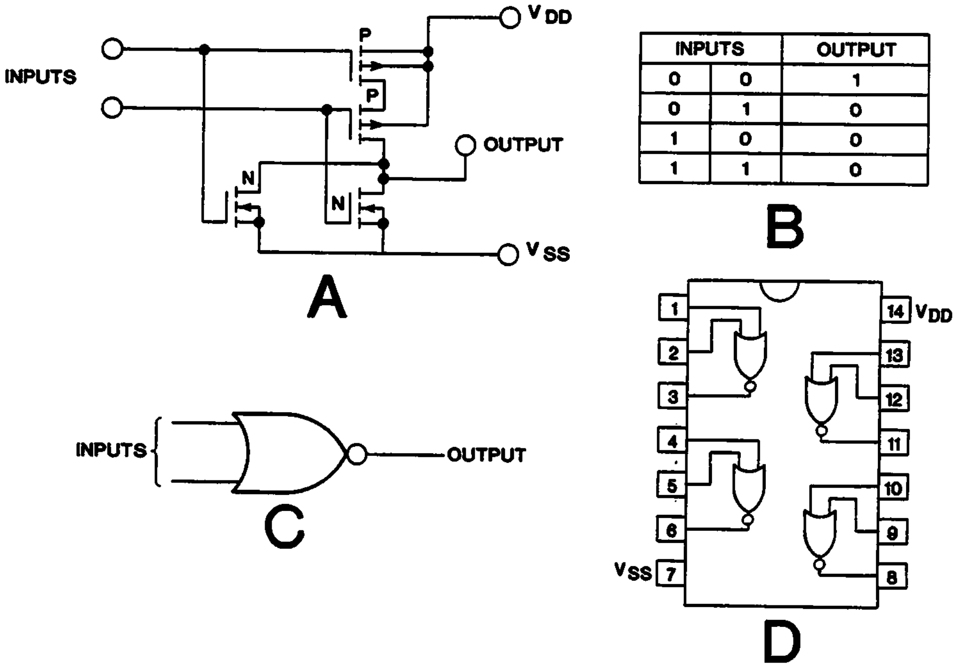
C



D

Figure 9-16

The circuit shown in Figure 9-17A contains four MOSFET's which are interconnected so that they can perform a useful logic function. The resulting circuit is referred to as a NOR gate, and like the NAND gate previously mentioned, it is a fundamental building block that is used to construct complex digital circuits. However, the NOR gate responds differently to various combinations of input voltage levels. The output levels (1 or 0) produced by the NOR gate for all possible combinations of input levels are shown in the table in Figure 9-17B.



| INPUTS | | OUTPUT |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

Figure 9-17
A CMOS digital IC.

The NOR gate symbol is shown in Figure 9-17C. This symbol is generally used in place of the actual schematic. Four of these NOR gates are usually formed on the one IC chip and mounted in a single IC package. An outline drawing of a typical dual in-line package, which contains four NOR gates, is shown in Figure 9-17D.

Both TTL and CMOS circuits can be used to perform NAND or NOR functions or a variety or other logic functions that must be performed in a highly complex digital system. The circuits shown in Figures 9-16 and 9-17 are therefore typical of the SSI circuits used in digital equipment. These circuits may be thought of as basic building blocks which can be used to construct complex digital systems that can perform useful operations.

# Linear IC's

As mentioned previously, linear circuits provide outputs that are proportional to their inputs. They do not switch between two states like digital circuits. The most popular linear circuits are the types that are designed to amplify dc and ac voltages. In fact, a high performance amplifier circuit known as an *operational amplifier*, is widely used in various types of electronic equipment.

The operational amplifier can amplify dc or ac voltages and has an extremely high gain. An operational amplifier can be constructed with discrete components, but it is more commonly produced in IC form and therefore sold as a complete package which is designed to meet certain specifications. However, the operational amplifier is designed so that it can be used in a variety of applications. You can control its gain by using additional external components and it usually has built-in features which make it possible to adjust its operation in various ways.

A typical operational amplifier circuit is shown in Figure 9-18A. The circuit contains transistors, resistors, and capacitors which are interconnected to form a highly efficient amplifying circuit. The circuit has two inputs and one output as shown. One inputis commonly referred to as the plus (+) or non-inverting input and the other is referred to as the minus (–) or inverting input. The circuit will amplify either dc or ac signals applied to either input. However, signals applied to the (+) input are not inverted when they appear at the output. In other words, as the input voltage goes positive or negative, the output voltage correspondingly goes positive or negative. When a signal is applied to the (–) input, inversion takes place. In other words, the polarity of the output signal is always opposite to that of the input signal. This unique feature greatly increases the versatility of the circuit.

When the input voltage is equal to zero, the output voltage should also be equal to zero. However, in practice the output voltage may be offset by a slight amount since component tolerance make it impossible to construct a perfectly balanced circuit. Therefore, two offset null terminals are provided so that the circuit can be appropriately balanced. This is done by simply connecting the opposite ends of a potentiometer to the offset null terminals, but the arm of the potentiometer is connected to circuit ground. The potentiometer may then be adjusted to balance the circuit.

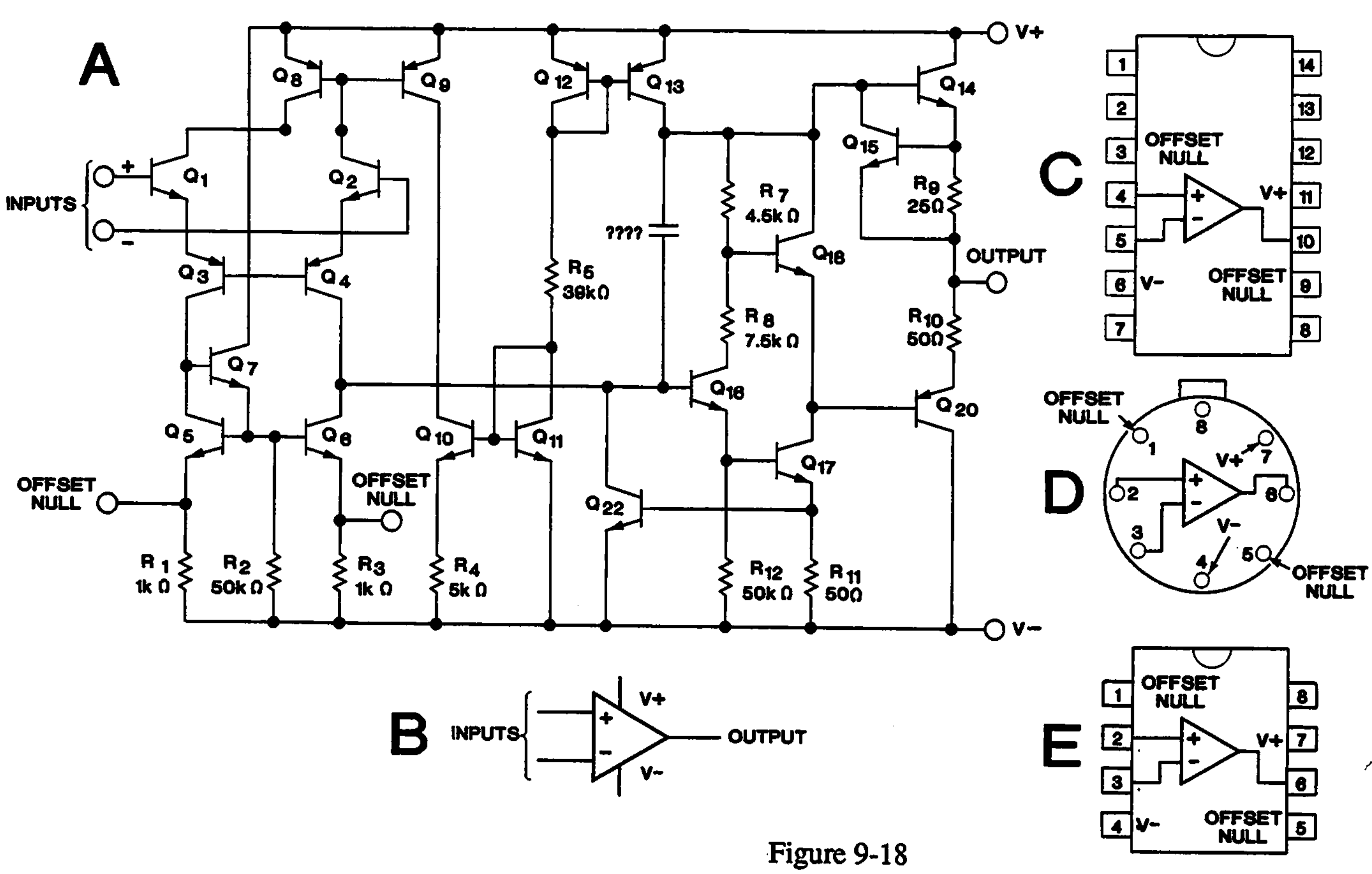


Figure 9-18
An operational amplifier: A typical linear IC.

Power is applied to the operational amplifier through terminals V+ and V–. The circuit therefore requires a positive voltage source and a negative voltage source. Most operational amplifiers can operate over a reasonably wide range of supply voltages, but these voltages should never exceed the maximum limits set by the manufacturer of the device. Operational amplifiers also consume very little power. Most units have maximum power dissipation ratings of 500 milliwatts or less.

The operational amplifier has an extremely high input resistance, but its output resistance is very low. The device also has an extremely high voltage gain. Many units are guaranteed to amplify an input voltage by at least 15,000 to 20,000 times and some of these units have typical gains that can exceed several hundred thousand or even more than a million.

In most cases, operational amplifiers are not used alone and their full amplifying capabilities are not utilized. Instead, it is common practice to connect external components to the amplifier in a manner which will allow a small portion of the output signal to return to the input and control the overall gain of the circuit. A lower gain is obtained in this manner, but circuit operation becomes more stable and predictable.

The operational amplifier is commonly represented by the symbol shown in Figure 9-18B. Notice that the (+) and (–) inputs are identified in the symbol. The operational amplifier circuit is generally packaged in a variety of ways to suit a broad range of applications. For example, the circuit in Figure 9-18A is available in a dual in-line package (DIP) as shown in Figure 9-18C or in a metal can package as shown in Figure 9-18D. It is even available as a mini DIP as shown in Figure 9-18E.

Operational amplifiers are used in various types of electronic equipment. They are the most important components used in electronic analog computers because their linear characteristics can be used to provide multiplying and summing operations. When used in analog computers, voltages are used to represent (are analogous to) the actual quantities to be multiplied or added. The extremely small operational amplifier IC's are suitable for use in portable electronic equipment where weight and power consumption must be held to a minimum. They are also suitable for use in portable test instruments and in communications equipment.

Various types of voltage regulator circuits are also constructed in IC form. These linear devices are used to convert an unregulated dc voltage (obtained through ac rectification) into a regulated dc output voltage which remains essentially constant while supplying a wide range of output currents. These voltage regulator IC's have replaced many of the discrete component regulators that were once widely used. Some IC regulators provide only one fixed output voltage, but other types are available which have adjustable outputs.

Special types of linear IC's are also designed for specific applications. For example, special linear IC's are designed for use in FM receivers where they are used to detect (recover the information in) FM signals. Some IC's are designed for use in solid-state color television receivers where they are used to detect, process and automatically control the chroma (color) signals. Others are used to provide extremely simple operations such as generating a signal to operate a lamp or some other indicator when an FM receiver or TV receiver is properly tuned. Some linear IC's are even used as an interface between digital circuits when digital information must be transmitted over a long transmission line. Such devices are commonly referred to as line drivers and receivers and they are used in digital systems even though they are basically linear devices.

# EXPERIMENT 11

## Operational Amplifier Characteristics

| | |
|---|---|
| *OBJECTIVE:* | *The purpose of this experiment is to demonstrate the operation of an IC op amp and to observe some of its characteristics.* |

### Introduction

In this experiment, you will become acquainted with some of the characteristics of an IC operational amplifier (op amp). An op amp is basically an amplifier circuit whose gain can be controlled very precisely with external discrete components.

### Material Required

Heathkit Analog Trainer
Voltmeter (VOM)
Oscilloscope
1 — IC Operational Amplifier 741 (442-22)
1 — 100 kilohm resistor (brown, black, yellow, gold)

## Procedure

1. Construct the circuit in Figure 9-19 following the wiring diagram in Figure 9-20. However, do not connect pins 7 and 4 to the positive and negative Trainer supplies at this time.

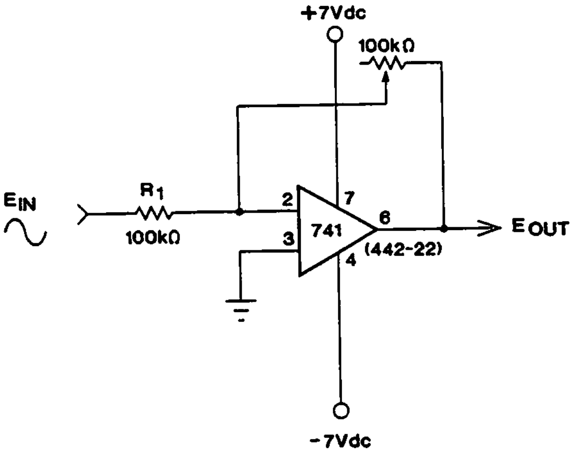


**Figure 9-19**

Experimental Circuit.

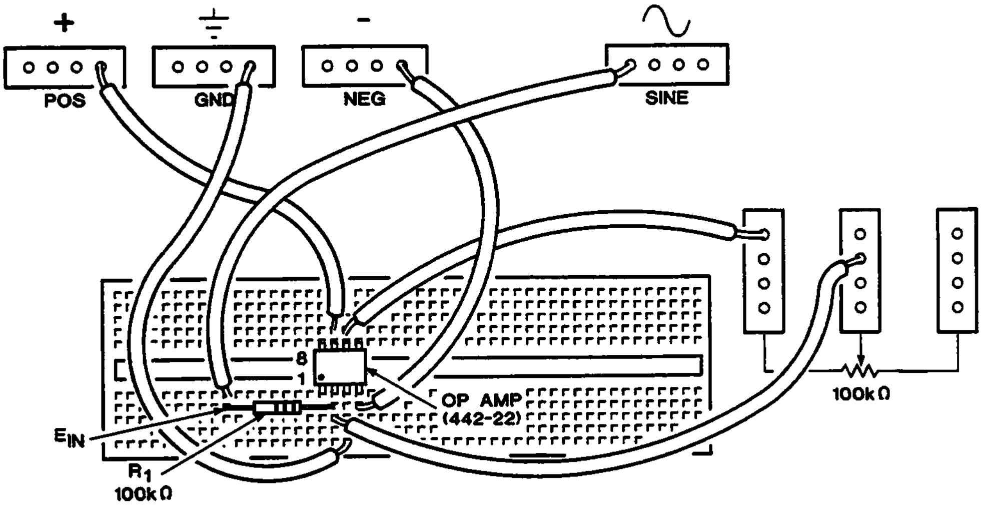


**Figure 9-20**

Wiring diagram for experimental circuit.

2. Apply power to the Trainer and connect your voltmeter to the POS and GND terminals of the power supply. Adjust the + voltage control for a +7 volts output.

3. Connect your voltmeter to the NEG and GND power supply terminals and adjust the – voltage control for a –7 volts output.

4. Connect pin 7 of the IC to the POS terminal and pin 4 to the NEG terminal.

5. Adjust the GENERATOR FREQ control for an output frequency of approximately 200 Hz.

6. Connect your oscilloscope to the input signal ($E_{in}$) of the circuit. Set your oscilloscope controls to display approximately four cycles at 1V/cm. Rotate the 100 kilohm potentiometer on the Trainer for maximum undistorted (unclipped) signal. Observe and record the peak-to-peak voltage waveform in Figure 9-21A.

7. If you have a dual trace oscilloscope, connect channel 2 to the IC output at pin 6. However, if you do not have a dual trace oscilloscope, move the test lead from $E_{in}$ to pin 6 of the IC. Do not disturb the oscilloscope controls unless it is absolutely necessary. Observe the output waveform and rotate the 100 kilohm control for the maximum undistorted output waveform. Record this waveform in Figure 9-21B. Record the peak-to-peak voltage waveform in the space provided in Figure 9-21B.

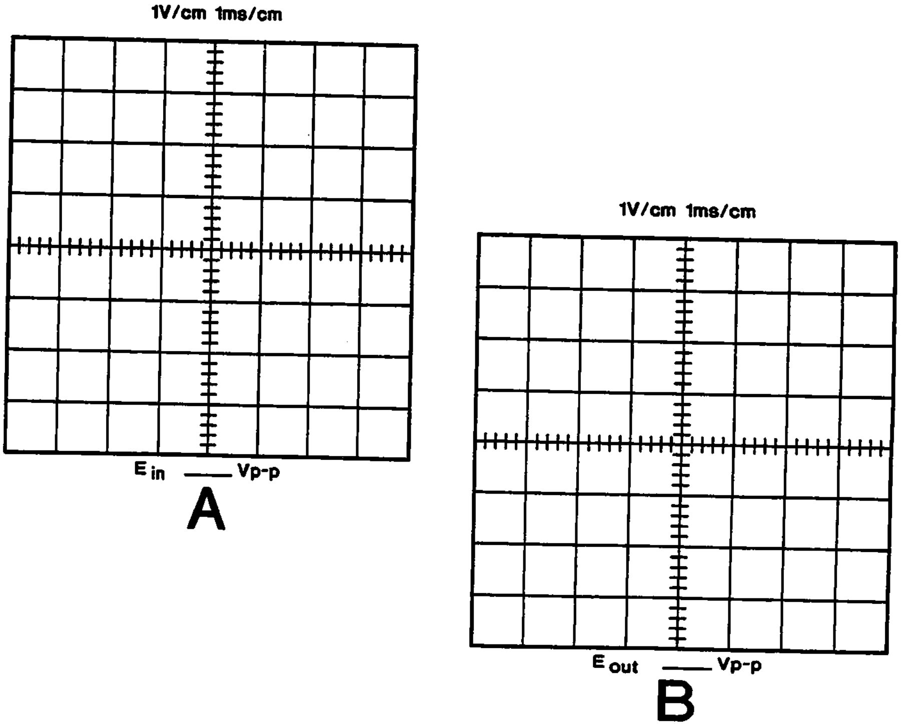


**Figure 9-21**

Chart for recording op amp waveforms.

Describe the phase relationship between the signal observed in step 6 and the one observed in step 7. ______________________________________________

______________________________________________

______________________________________________

8. Calculate the gain of the op amp, using the formula:

$$A_V = E_{out}/E_{in}$$

$A_V =$ ______

9 This completes the experiment. Turn off your Trainer.

## Discussion

In this experiment, you observed some of the characteristics of a typical operational amplifier. Op amps are extremely versatile linear IC's that have a large number of applications in electronics. One advantage of op amps is that they are capable of providing high gain over a wide range of frequencies and the amount of gain can be controlled very closely. One method of controlling gain is by means of negative feedback. The circuit in this experiment uses negative feedback to control gain.

Another advantage of op amps is that they can be used as either inverting or noninverting amplifiers. When an op amp is used as an inverting amplifier, as in this experiment, the output signal is 180 degrees out of phase (inverted) with the input signal. If the IC is used as a noninverting amplifier, the output signal is in phase with the input signal.

## Self Test Review

31. Integrated circuits may be broadly classified as either ____________ or ____________ IC's.

32. The IC shown in Figure 9-16A is commonly referred to as a ____________ gate.

33. Like all digital circuits, the circuit in Figure 9-16A responds to only ____________ input voltage levels.

34. The voltage levels used in a digital system are commonly referred to as ____________ and ____________ or ____________ and ____________ logic levels.

35. Power must be applied to the IC package shown in Figure 9-16D through pins ____________ and ____________.

36. The digital IC in Figure 9-17A utilizes both P- and N-channel ____________.

37. The circuit in Figure 9-17A is commonly called a ____________ gate.

38. A CMOS digital IC usually has a higher input ____________ than a TTL digital IC.

39. A linear circuit provides an output that is ____________ to its input.

40. One of the most popular linear IC's is the ____________ amplifier.

41. When an ac signal is applied to the (–) input of the circuit shown in Figure 9-18A, it is ____________ when it appears at the output.

42. The IC package outlined in Figures 9-18E is commonly referred to as a ____________.

43. Positive and negative voltage sources must be applied to the IC package shown in Figure 9-18C through pins ____________ and ____________ respectively.

### Answers

31. digital or linear
32. NAND
33. two
34. high and low or 1 and 0
35. 14 and 7
36. MOSFET's
37. NOR
38. resistance
39. proportional
40. operational
41. inverted
42. mini DIP
43. 11 and 6

# UNIT SUMMARY

The integrated circuit has revolutionized the electronics industry because it has brought about tremendous reductions in the size and weight of electronic equipment. The use of IC's has also resulted in greater equipment reliability and lower operating expense. IC's are now used extensively throughout the electronics industry and new applications for these devices are constantly being found.

Integrated circuits are constructed in basically four different ways. They may be produced by using monolithic, thin-film, thick-film or hybrid techniques. Monolithic IC's are the most widely used devices. Monolithic circuits are formed within a single piece of semiconductor material known as a substrate. The components most commonly found in these circuits are diodes, transistors, resistors, and capacitors.

Monolithic circuits can be formed by using bipolar or MOS construction techniques. Bipolar IC's utilize the same basic technology associated with the construction of bipolar transistors. MOS IC's are produced with the same basic technology that is used to construct MOS field effect transistors.

Thin-film and thick-film circuits are formed by depositing layers of metals and oxides on insulating substrates. The components that are formed in a thin-film circuit are so thin that they have almost no discernable thickness. The components formed in thick-film circuits are much larger and more closely resemble discrete components in appearance. Film techniques are used to construct resistors, capacitors, and interconnecting conductors, but components such as diodes and transistors must be formed separately out of semiconductor materials and attached to the film circuit.

Hybrid circuits are formed by combining monolithic, thin-film, and thick-film construction techniques. The hybrid approach to the construction of IC's makes it possible to utilize the best features found in each type of IC.

Integrated circuits can be packaged in a variety of ways. In most cases, the package (no matter how small) is considerably larger than the IC itself. The package protects the IC from various contaminants which could affect its operation and at the same time provides suitable mounting pins or leads which can be easily inserted in a matching socket or receptacle or soldered in place. The DIP, mini DIP, flat-pack and metal can packages are among the types that are most commonly used. However, the advent of the new SMT packages has brought about a revolution in circuit board design and construction.

All integrated circuits, no matter how they are constructed, can be classified as either digital or linear devices. Digital IC's contain switching circuits which are capable of performing various logic functions. Linear IC's operate continuously and do not switch from one state to another. Linear circuits provide an output that is proportional to their respective inputs.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers with the correct ones that appear on the Examination Answer Sheet which follows.

1. The four components that are most commonly used in an IC are

   A. diodes, resistors, inductors, and capacitors.
   B. diodes, transistors, resistors, and inductors.
   C. diodes, transistors, resistors, and capacitors.
   D. transistors, resistors, inductors, and capacitors.

2. When compared to a discrete component circuit, the IC is

   A. smaller, more reliable, and often less expensive.
   B. smaller, less reliable, and less expensive.
   C. smaller, more reliable, and generally more expensive.
   D. larger, more liable, and less expensive.

3. The four basic types of integrated circuits are the

   A. silicon, germanium, P-type, and N-type IC's.
   B. monolithic IC's, bipolar IC's MOS Ic's, and unipolar IC's.
   C. monolithic IC's, thin-film circuits, thick-film circuits, and hybrid circuits.
   D. monolithic, film-type, hybrid, and bipolar IC's.

4. Many monolithic IC's are simultaneously formed on a

   A. glass substrate.
   B. ceramic substrate.
   C. circular ceramic wafer.
   D. circular semiconductor wafer.

5. Monolithic IC's may be formed by utilizing the construction techniques associated with bipolar transistors or field effect transistors and are usually classified as

   A. bipolar Ic's or MOS IC's.
   B. bipolar IC's or FET IC's.
   C. bipolar Ic's or unipolar IC's.
   D. MOS IC's or FET IC's.

6. If a monolithic IC has only 20 components, it would be classified as an

   A. LSI circuit.
   B. MSI circuit.
   C. SSI circuit.
   D. MOS circuit.

7. Integrated circuits that are formed by depositing layers of metals and oxides on insulating substrates by using evaporation or sputtering techniques, are referred to as

   A. MOS circuits.
   B. thin-film circuits.
   C. thick-film circuits.
   D. hybrid circuits.

8. When silk-screening techniques are used to form components and conductors on an insulating substrate, the resulting circuit is referred to as a

   A. MOS circuit
   B. bipolar circuit.
   C. thin-film circuit.
   D. thick-film circuit.

9. Monolithic IC's, thin-film and thick-film circuits, and also discrete components can be combined in various arrangements to produce a

   A. bipolar Ic.
   B. hybrid circuit.
   C. MOS circuit.
   D. LSI circuit.

10. The DIP is a popular type of

    A. monolithic IC package.
    B. hybrid IC package.
    C. thick-film IC package.
    D. transistors package that is used with IC's.

11. Digital IC's are basically switching circuits which respond to

    A. a continuously changing dc voltage.
    B. only zero volt and +3 volt levels.
    C. high and low logic levels.
    D. voltages that are between zero and +3 volts.

12. Two linear IC's that are commonly used are the

    A. TTL NAND gate and CMOS NOR gate.
    B. operational amplifier and TTL NAND gate.
    C. operational amplifier and CMOS NOR gate.
    D. operational amplifier and voltage regulator.

# EXAMINATION ANSWERS

1. C — Diodes and transistors are easier to construct and therefore more popular than resistors and capacitors, but all four of these components are used in IC's. Inductors cannot be made by integrated circuit techniques.

2. A — These are the most important advantages that the IC has to offer, although there are many other fringe benefits to be obtained from using them.

3. C — Most of the IC's presently manufactured can be classified as one of these four basic types.

4. D — The silicon wafer is usually less than two inches in diameter and may contain hundreds of IC's.

5. A — Bipolar IC's are formed using bipolar transistor construction techniques associated with MOS field effect transistors.

6. C — An SSI (small scale integration) circuit usually has less than 100 components.

7. B — In thin-film circuits, resistors and capacitors are formed by depositing layers of metals and oxides. If diodes and transistors are required, they must be formed separately and attached to the thin-film circuit.

8. D — The silk-screening process produces components (resistors and capacitors) which are much thicker and larger than thin-film components. In these thick-film circuits, diodes and transistors must still be produced as separate components and then added to the circuits.

9. B — The hybrid approach to circuit construction makes it possible to combine the best features found in each of the basic types of integrated circuits.

10. A — Both large DIP's (dual in-line packages) and smaller mini DIP's are widely used with monolithic circuits. These packages are generally made of plastic or ceramic materials.

11. C — Instead of referring to the actually voltage levels involved, the terms high and low or 1 and 0 are used.

12. D — Operational amplifiers and voltage regulators are just two of the many types of linear IC's that are commonly used.

## *Unit 10*

# OPTOELECTRONIC DEVICES

# CONTENTS

Introduction .Page 10-3

Unit Objectives .Page 10-4

Unit Activity Guide .Page 10-5

Basic Principles of Light .Page 10-6

  Characteristics .Page 10-6

  Units of Measurement .Page 10-9

    Radiometric System .Page 10-10

    Photometric System .Page 10-11

Light-Sensitive Devices .Page 10-16

  Photoconductive Cells .Page 10-16

  Photovoltaic Cells .Page 10-19

  Photodiodes .Page 10-22

  Phototransistors .Page 10-26

Experiment 12 .Page 10-33

Light-Emitting Devices .Page 10-39

  LED Operation .Page 10-40

  LED Construction .Page 10-41

  LED Characteristics .Page 10-43

  LED Applications .Page 10-47

Liquid Crystal Displays .Page 10-51

Experiment 13 .Page 10-56

Unit Summary .Page 10-61

Unit Examination .Page 10-62

Examination Answers .Page 10-65

# INTRODUCTION

In this unit, you will examine a group of solid-state components that are capable of converting light energy into electrical energy or electrical energy into light energy. These components are commonly referred to as *optoelectronic* devices, since their operation realize on both optic and electronic principles.

The optoelectronic devices described in this unit are divided into three basic groups. They are classified as *light-sensitive* devices, *light-emitting* devices and light transmissive or reflective devices. The most important components found within these two categories are described in detail and the various ways in which these components can be used are considered.

Like other fields, optoelectronics has its own peculiar set of terms, definitions, and units of measurement. Although it is not necessary to understand all of the terms that are used, you must at least learn certain portions of the optoelectronics language if you are to fully appreciate and understand the operation of the various optoelectronic components. The important concepts that are involved are, therefore, developed at the beginning of this unit.

Study this unit carefully. A familiarity with optoelectronic devices is becoming increasingly important since these components are now used extensively, and new applications are constantly being found. Also, the optoelectronic theory that you learn in this unit will greatly enhance your knowledge of semiconductor devices.

# UNIT OBJECTIVES

When you have completed this unit on optoelectronic devices you should be able to:

1. Describe the major characteristics of light.
2. Determine the wave length of any given light frequency.
3. Explain the difference between the radiometric and photometric systems for measuring light.
4. Describe the basic function of a light-sensitive device.
5. Name four photosensitive devices and briefly describe their operation.
6. Describe the basic function of a light-emitting device.
7. Describe the basic principle of operation of the light-emitting diode.
8. Name three advantages that the light-emitting diode has over an incandescent or neon lamp.
9. Explain the basic principles of operation of a liquid-crystal display.
10. Identify at least one advantage of liquid-crystal over other types of displays as well as at least one disadvantage of liquid crystal displays.

# UNIT ACTIVITY GUIDE

| | **Completion Time** |
|---|---|
| ☐ Read "Basic Principles of Light." | ________ |
| ☐ Answer Self Test Review Questions 1-14. | ________ |
| ☐ Read "Light-Sensitive Devices." | ________ |
| ☐ Answer Self Test Review Questions 15-27. | ________ |
| ☐ Perform Experiment 12 on "Phototransistor Characteristics." | ________ |
| ☐ Read "Light-Emitting Devices." | ________ |
| ☐ Read "Liquid-Crystal Displays." | ________ |
| ☐ Answer Self Test Review Questions 28-41. | ________ |
| ☐ Perform Experiment 13 on "LED Characteristics." | ________ |
| ☐ Read the Unit Summary. | ________ |
| ☐ Complete the Unit Examination. | ________ |
| ☐ Check the Examination Answers. | ________ |

# BASIS PRINCIPLES OF LIGHT

In order to understand the operation of optoelectronic devices it is necessary to understand the basic principles of light. You must know what light is and you must be familiar with some of its basic units of measurement. Therefore, we will begin this unit by defining light and considering its various properties. Then we will briefly examine the basic techniques which are used to measure light. This initial discussion will include only the most basic and important terms and definitions.

## Characteristics

The term *light* is used to identify *electromagnetic radiation* which is visible to the human eye. Basically, light is just one type of electromagnetic radiation and differs from other types such as cosmic rays, gamma rays, X-rays, and radio waves only because of its frequency.

The light spectrum extends from approximately 300 gigahertz* to 300,000,000 gigahertz. It is wedged midway between the high end of the radio frequency (RF) waves, which roughly extend up to 300 gigahertz, and the X-rays, which begin at roughly 300,000,000 gigahertz. Above the X-ray region are the gamma rays and then the cosmic rays.

Like other types of electromagnetic radiation, light propagates (travels) through space and certain types of matter. The movement of light through space can be compared to the movement of radio waves which periodically fluctuate in intensity as they move outward from an antenna or some other radiating body. Therefore, light waves can be measured in terms of **wavelength**. As with radio waves, the wavelength of light waves is determined by the velocity at which light is moving and the frequency of its fluctuations. This relationship can be summed up with the following mathematical equation:

$$\lambda = \frac{v}{f}$$

---

*One gigahertz, abbreviated GHz, is equal to a frequency of 1,000,000,000 Hertz or $10^9$ cycles per second.

Where λ (the Greek letter lambda) represents one complete wavelength, the letter v represents velocity and f represents frequency. This equation states that one complete wavelength is equal to the velocity divided by the frequency.

Light travels at an extremely high velocity. In a vacuum its velocity is 186,000 miles per second or $3 \times 10^{10}$ (30,000,000,000) centimeters per second. Its velocity is only slightly lower in air but when it passes through certain types of matter, such as glass or water, its velocity is reduced considerably.

When we substitute the velocity of light in a vacuum (in centimeters per second) into the equation given above along with its frequency in hertz, we obtain the wavelength in centimeters (cm). Since light has a frequency range which extends from 300 gigahertz ($300 \times 10^9$ hertz) to 300,000,000 gigahertz ($300 \times 10^{15}$ hertz) its wavelength will vary over a considerable range. The minimum wavelength occurs at the maximum frequency and is equal to:

$$\lambda = \frac{3 \times 10^{10}}{300 \times 10^{15}} = 0.0000001 \text{ centimeters}$$

The maximum wavelength occurs at the minimum frequency and is equal to:

$$\lambda = \frac{3 \times 10^{10}}{300 \times 10^{9}} = 0.1 \text{ centimeters}$$

A frequency range of 300 to 300,000,000 gigahertz therefore corresponds to wavelengths which vary from 0.1 to 0.0000001 centimeters. Since the wavelengths associated with the various light frequencies are extremely short, it is common practice to express them in much smaller units. For example, 0.0000001 centimeters ($0.0000001 \times 10^{-2}$ meters) may be expressed as $0.001 \times 10^{-6}$ meters or 0.001 micrometers (0.001 $\mu$m). However, it may also be expressed as simply 0.001 microns. Also, 0.0000001 centimeters is equal to $1 \times 10^{-9}$ meters or 1 nanometer (1 nm). Another unit that is commonly used is the *angstrom* which is represented by the symbol Å. One angstrom equals $1 \times 10^{-10}$ meters. Therefore, 0.0000001 centimeters (which is equal to $10 \times 10^{-10}$ meters) is equal to 10 angstroms.

It is important to understand the relationship between frequency and wavelength when dealing with either radio waves or light waves. These two terms are often used interchangeably in any discussion which concerns the utilization of electromagnetic radiation. You will see both of these terms used as you proceed through this unit.

Within the 300 to 300,000,000 GHz light spectrum only a narrow band of frequencies can actually be detected by the human eye. This narrow band of frequencies appears as various colors such as red, orange, yellow, green, blue and violet. Each color corresponds to a very narrow range of frequencies within the visible region. The entire visible region extends from slightly more than 400,000 GHz to approximately 750,000 GHz. Above the visible region (between 750,000 and 300,000,000 GHz) the light waves cannot be seen. The light waves which fall within this region are referred to as *ultraviolet rays*. Below the visible region (between 300 and 400,000 GHz) the light waves again cannot be seen. The light waves within this region are commonly referred to as *infrared rays*. The entire light spectrum is shown in Figure 10-1 so that you can compare the three regions just described.

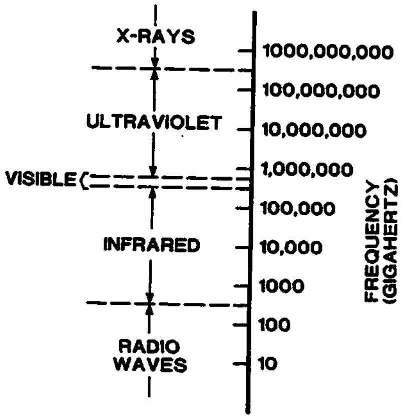


Figure 10-1
The light spectrum

Although light is assumed to propagate as electromagnetic waves, the wave theory alone cannot completely explain all of the phenomena associated with light. For example, the wave theory may be used to explain why light bends when it flows through water or glass. However, it cannot explain the action that takes place when light strikes certain types of semiconductor materials, and it is this resultant action that forms the basis for much of the optoelectronic theory presented in this unit. In order to explain why and how semiconductor materials are affected by light it is necessary to assume that light has additional characteristics.

To adequately explain the operation of the optoelectronic devices included in this unit it is necessary to consider an additional aspect of light as explained by basic *quantum theory*. Quantum theory acknowledges that light has wave-like characteristics but it also states that a light wave behaves as if it consisted of many tiny particles. Each of these tiny particles represents a discrete quanta or packet of energy and is called a *photon*.

The photons within a light wave are uncharged particles and their energy content is determined by the frequency and wavelength of the wave. The higher the frequency, the more energy each photon will contain. This means that the light waves at the upper end of the light spectrum possess more energy than the ones at the lower end of the spectrum. This same rule also applies to other types of electromagnetic radiation. For example, X-rays have a higher energy content than light waves while light waves possess more energy than radio waves.

Therefore, light has a dual personality. It propagates through space like radio waves, but it behaves as if it contains many tiny particles. This particle-like property of light will be utilized in this unit to explain the action that takes place in various types of optoelectronic components.

## Unit of Measurement

In your previous studies you learned how current, voltage and resistance are related. But even more important was the fact that you learned how to measure these electrical quantities. It was necessary to create suitable units of measurement for each of these quantities so that their precise values could be determined. The same is true when dealing with light. We must know how to accurately measure the various properties of light so that it can be effectively utilized. However, when dealing with light, several problems arise which are not encountered when dealing with the electrical quantities mentioned above. First of all, there is not one but two systems in common use for measuring light. Furthermore, these two systems are similar but not identical and it is often difficult to convert from one system to the other.

One system of light measurement is based on the response of the human eye and is applicable only to that portion of the light spectrum that is visible. This method of describing or quantifying light is referred to as the *photometric* system. The second system is used to measure both visible and nonvisible light. Therefore, it is applicable throughout the entire optical spectrum. This method is referred to as the *radiometric* system.

Since the photometric and radiometric systems are both used in the optoelectronics field, we will briefly examine both of these systems. However, at this time we will consider only the most basic and important terms and units of measurement that are involved in each system. The less important terms will be described as required later in the unit.

## Radiometric System

In the radiometric system the basic units of measurement are the watt and the centimeter and all of the additional units within this system are derived from these. A number of radiometric terms are required to adequately define and measure the various aspects of the light that is emitted from a given source or the light that strikes the surface of an object. Furthermore the names of these various radiometric quantities always have the prefix *radiant* before them. With these thoughts in mind, we will not briefly review the more important radiometric terms and units of measurement.

In radiometric terminology the energy traveling from a light source is considered to be in the form of electromagnetic waves and is referred to as *radiant energy*. Radiant energy is often represented by the symbol $Q_e$.

The total amount of radiant energy supplied by a light source per unit time, or in other words, the total rate of flow of the radiant energy is commonly referred to as *radiant flux* or *radiant power*. Radiant flux is commonly represented by the symbol $\Phi_e$ and is measured in joules per second or *watts*.

The term *radiant intensity* may also be used to describe the amount of light produced by a source. However, this term is more specific than the one previously defined and can be used to describe the distribution of radiant energy. Radiant intensity is simply the radiant flux per unit solid angle that is traveling in a given direction. It is represented by the symbol $I_e$, and it is measured in *watts per steradian*. The steradian is a dimensionless unit much like the radian. It is simply a unit of solid angular measurement. One steradian of solid angle which has its apex at the center of a sphere that has a radius of 1 meter will subtend an area of 1 square meter on the sphere's surface. Any given sphere will contain a total of $4\pi$ steradians.

The two previous terms are useful in describing the radiant energy emitted from a point source of light which radiates in all directions. However, additional terms are needed to describe the radiant energy that is either emitted from or striking a specified surface area. For example, the radiant energy striking a surface per unit area is referred to as *irradiance* or *radiant incidence* and is represented by the symbol $E_e$. It is usually measured in *watts per square meter*. However, the radiant energy that is emitted from or leaving a surface per unit area is referred to as *radiant exitance*. Radiant exitance is represented by the symbol $M_e$ and is also measured in *watts per square meter*.

Another radiometric term that is widely used is *radiance*. Like the previous term radiant exitance, this term applies to radiant energy that is spread over a specified area. Radiance is defined as the radiant intensity per unit area that is either leaving, passing through, or arriving at a surface in a certain direction. However, in this case the surface area is the area as viewed from the specified direction. Radiance is represented by the symbol $L_e$ and is measured in *watts per steradian per square meter*. For a wide area light source, the radiance remains constant at all viewing angles with respect to the source. This is because the radiant intensity, which is greatest when the surface area is viewed from directly above, decreases proportionally with the apparent area of the source as the viewing angle (from directly above) increased.

Additional radiometric terms exist but the terms just described are the ones that are most commonly used. It is not necessary to remember all of these terms at this time. Their full meaning cannot be realized until they are used in various practical applications.

## Photometric System

The various concepts that were just described using radiometric terminology can also be described using photometric terms. However, the photometric terms always have the prefix *luminous* before them. Furthermore, the photometric terms are not exactly equivalent to their corresponding radiometric terms. The measurements in photometry are limited to the visible light waves while radiometry deals with the entire light spectrum as mentioned earlier.

Any photometric light measuring device must respond in much the same manner as the human eye. However, the response of the human eye is slightly different at normal and low light levels. At low light levels the visible light frequencies are shifted to a slightly lower range and the eye is said to have a *scotopic* response. At normal light levels the eye responds to a slightly higher range of frequencies and is said to have a *photopic* response. Unless otherwise specified, the various measurements in the photometric system are based on the photopic response of the eye. With these thoughts in mind we will now examine the various photometric terms and units of measurement which are related to the radiometric terms previously described.

In the photometric system the light energy produced by a given source is referred to as *luminous energy*. Luminous energy is represented by the symbol $Q_v$. The amount of luminous energy produced by a source per unit time is referred to as *luminous flux* or *luminous power*. Luminous flux is represented by the symbol $\Phi_v$ and is measured in *lumens*. The lumen is considered to be the basic unit of measurement in the photometric system and it may be compared to the watt in the radiometric system. It takes 680 lumens to equal 1 watt of radiant flux, but this is true only for light energy for which the human eye is most sensitive. The conversion is therefore valid only for a specific green light which has a wavelength of exactly 0.555 micrometers ($0.555 \times 10^{-6}$ meters). At any other wavelength, the human eye is less sensitive and there are fewer lumens to each watt. Herein lies the difficulty of converting from radiometric units to photometric units and vice-versa. To make accurate conversions at all wavelengths would require consideration of the spectral response of the human eye and the spectral output of a given light source. Such conversions can be cumbersome and time consuming and will not be considered here.

The luminous flux per unit solid angle traveling in a specific direction from a light source is referred to as *luminous intensity*. This term may be compared to the radiometric term radiant intensity. Luminous intensity is represented by the symbol $I_v$ and its unit of measurement is the *candela*. One candela is equal to 1 lumen of luminous flux per steradian.

The luminous flux striking a surface per unit area is referred to as *illuminance* or *illumination* and is represented by the symbol $E_v$. The popular unit of measurement for illuminance is the *lux*. One lux is equal to 1 lumen per square meter. However, other units of measurement are sometimes used. For example, illuminance may be expressed in lumens per square foot or *footcandles*. One footcandle equals 10.76 lux.

The luminous flux that is emitted from a unit area of a surface is referred to as *luminous exitance*. Luminous exitance is represented by the symbol $M_v$ and is measured in *lumens per square meter*. This term was once known as luminous emittance but this older term is now seldom used.

The last photometric term that we will identify is *luminance*. Luminance is simply the luminance intensity per unit area that is either leaving, passing through, or arriving at a surface in a specific direction. The surface area is always the apparent area as viewed from the specified direction. Luminance is represented by the symbol $I_L$ and is measured in *lumens per steradian per square meter* or *candela pre square meter*. One candela per square meter is sometimes called a 1 *nit* of luminance. Other units are also used which are based on the candela over various unit of area but they will not be considered in the brief discussion.

The photometric term luminance may be compared to the radiometric term radiance, although the two terms are not exactly equivalent. Furthermore, the term luminace may be thought of as the photometric equivalent of the term *brightness*.

As in the radiometeric system, additional photometric terms exist. However, the terms just described are the most important ones and they are related to the radiometric terms that were described earlier.

# Self Test Review

1. Light is considered to be ____________ radiation that is visible to the human eye.

2. The light spectrum extend from ____________ gigahertz to ____________ gigahertz.

3. In terms of wavelength the light spectrum extends from ____________ to ____________ centimeters.

4. Most of the frequencies within the light spectrum are visible.

   a. True
   b. False

5. The invisible frequencies within the light spectrum are referred to as __________ and __________ rays.

6. Light has wave-like characteristics but it also behave as if it consisted of many tiny particles known as ____________.

7. The ___________ system of measuring light is based on the response of the human eye.

8. The ___________ system of measuring light is applicable through-out the entire light spectrum.

9. Radiometric terms always use the prefix ___________.

10. Photometric terms always used the prefix ___________.

11. The radiometric terms ______ ______ and ________________ are useful in describing the radiant energy emitted from a point source.

12. The radiometric terms ______ ______, ____________ and ____________ are used to describe radiant energy that is spread over a specified areas.

13. For each radiometric term there is a corresponding photometric term which describes the same basic concept.

    a. True
    b. False

14. The corresponding radiometric and photometric terms are exactly equivalent.

    a. True
    b. False

# Answers

1. electromagnetic
2. 300 gigahertz to 300,000,000 gigahertz
3. 0.1 to 0.000001 centimeters
4. b. False — only a small portion of the light spectrum is visible.
5. ultraviolet and infrared
6. photons
7. photometric
8. radiometric
9. radiant
10. luminous
11. radiant flux (or radiant power) and radiant intensity.
12. irradiance (or radiant incidence), radiant exitance and radiance.
13. a. True
14. b. False — The radiometric system describes light at all frequencies while the photometric system describes only visible light thus making conversions from one system to the other difficult and time consuming.

# LIGHT SENSITIVE DEVICES

Most of the optoelectronic devices that are now used are basically light-sensitive devices. In other words they simply respond to changes in light intensity by either changing their internal resistance or by generating a corresponding output voltage. We will now examine some of the most important light-sensitive devices.

## Photoconductive Cells

The *photoconductive cell* is one of the oldest optoelectronic components. Basically it is nothing more than a light-sensitive resistor whose internal resistance changes as the light shining on it changes in intensity. The resistance of the device decreases nonlinearly with an increase in light intensity. That is, the resistance decreases, but the decrease is not exactly proportional to the increase in light.

Photoconductive cells are usually made from light-sensitive materials such as cadmium sulfide (Cd S) or cadmium selenide (Cd Se), although other materials such as lead sulfide and lead telluride have been used. These basic materials may also be doped with other materials such as copper or chlorine to control the exact manner in which the resistance of the device varies with light intensity.

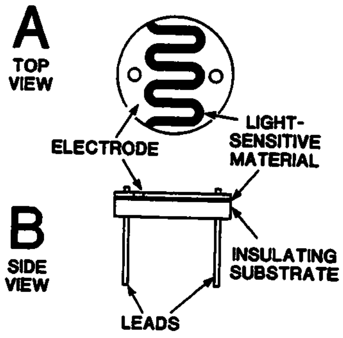


Figure 10-2
A typical photoconductive cell. (A) topview (B) sideview

Figure 10-2 shows how a typical photoconductive cell is constructed. A thin layer of light-sensitive material is formed on an insulating substrate which is usually made from glass or ceramic materials. Then two metal electrodes are deposited on the light-sensitive material as thin layers. The top view (Figure 10-2A) shows that the electrodes do not touch but leave an S-shaped portion of the light-sensitive material exposed. This allows greater contact length but at the same time confines the light-sensitive material to a relatively small area between the electrodes. Two leads are also inserted through the substrate and soldered to the electrodes as shown in the side view in Figure 10-2B. The photoconductive cell is often mounted in a metal or plastic case (not shown) which has a glass window that will allow light to strike the light-sensitive material. Also, the electrodes used with some cells may be arranged in more complicated patterns and the entire cells may be quite large (1 or more inches in diameter) or relatively small (less than 0.25 inches in diameter).

Photoconductive cells are more sensitive to light than other types of light-sensitive devices. The resistance of a typical cell might be as high as several hundred megohms when the light striking its surface (its illumination) is zero (complete darkness) and as low as several hundred ohms when the illumination is over 100 lux (approximately 9 footcandles). This represents a tremendous change in resistance for a relatively small change in illumination. This extreme sensitivity makes the photoconductive cell suitable for applications where light levels are low and where the changes in light intensity are small. However, these devices do have certain disadvantages. Their greatest disadvantage is the fact that they respond slowly to changes in illumination. In fact they have the slowest response of all light-sensitive devices. Also, they have a light memory or history effect. In other words, when the light level changes, the cell tends to remember previous illumination. The resistance of the cell at a specific light level is a function of the intensity, the duration of its previous exposure, and the length of time since that exposure.

Most photoconductive cells can withstand relatively high operating voltages. Typical devices will have maximum voltage ratings of 100, 200, or 300 volts dc. However, the maximum power consumption for these devices is relatively low. Maximum power ratings of 30 milliwatts to 300 milliwatts are typical.

The photoconductive cell is often represented by one of the schematic symbols shown in Figure 10-3. The symbol in Figure 10-3A simply consists of a resistor symbol inside of a circle. Two arrows are also used to show that the device is light-sensitive. The symbol in Figure 10-3B is similar but it contains the Greek letter λ (lambda) which is commonly used to represent the wavelength of light.

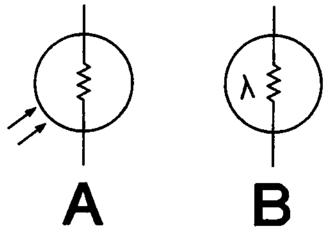


Figure 10-3
Commonly used photoconductive cell symbols.

Photoconductive cells have many applications in electronics. For example, they are often used in devices such as intrusion detectors and automatic door openers where it is necessary to sense the presence or absence of light. However, they may also be used in precision test instruments which can measure the intensity of light. A simple intrusion detector circuit is shown in Figure 10-4. The light source projects a narrow beam of light onto the cell and this causes the cell to exhibit a relatively low resistance. The cell is in series with a sensitive ac relay and its 120 volt ac, 60 Hz power source. The cell allows sufficient current to flow through the circuit and energize the relay. When an intruder breaks the light beam, the cell's resistance increases considerably and the relay is deactivated. At this time the relay contacts close and apply power (from a separate dc source) to an alarm, which sounds a warning. A relay is used because it is capable of controlling the relatively high current that is needed to operate the alarm. When a large relay is used, the photoconductive cell (because of its low power or current rating) may not be able to directly control the relay. In such a case the photoconductive cell is used to control a suitable amplifier circuit which in turn generates enough current to drive the relay.

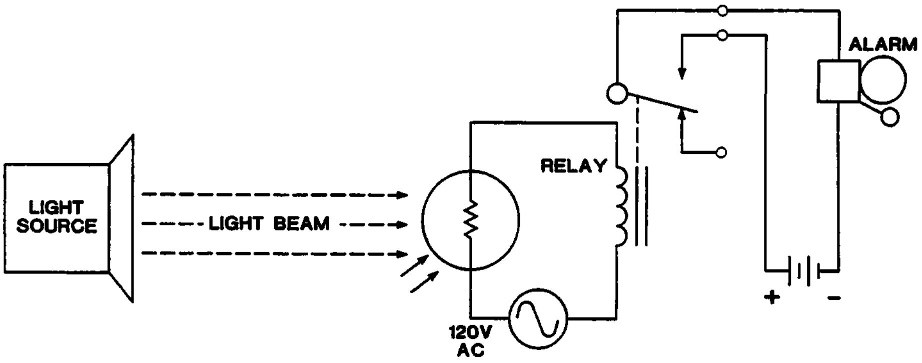


Figure 10-4
A basic intrusion detector circuit.

Since the photoconductive cell is constructed from a bulk material and does not have a PN junction, it is a bidirectional device. In other words it exhibits the same resistance in either direction and may therefore be used to control either dc or ac. Due to its bulk construction, the photoconductive cell is often referred to as a *bulk photoconductor*. However, you may also see it referred to as a *photoresistive cell* or simply a *photocell*.

# Photovoltaic Cells

The *photovoltaic cell* is a device which directly converts light energy into electrical energy. When exposed to light this device generates a voltage across its terminals and this voltage increases as the light increases in intensity. The photovoltaic cell has been used for a number of years in various military and space applications. It is commonly used aboard satellites and spacecraft to convert solar energy into electrical power which can be used to operate various types of electronic equipment. Since most of its applications generally involve the conversion of solar energy into electrical energy, this device is commonly referred to as a *solar cell.*

The photovoltaic cell is basically a junction device which is made from semiconductor materials. Although many different semiconductor materials have been used, the device is usually made from either silicon or selenium with silicon being the preferred material. The basic structure of a silicon photovoltaic cell is shown in Figure 10-5. Notice that the device has a P-type layer and an N-type layer which form a PN junction and a metal backplate or support is placed against the N-type layer. Also, a metal ring is attached to the outer edge of the P-type layer. These pieces of metal serve as electrical contacts to which the two external leads may be attached.

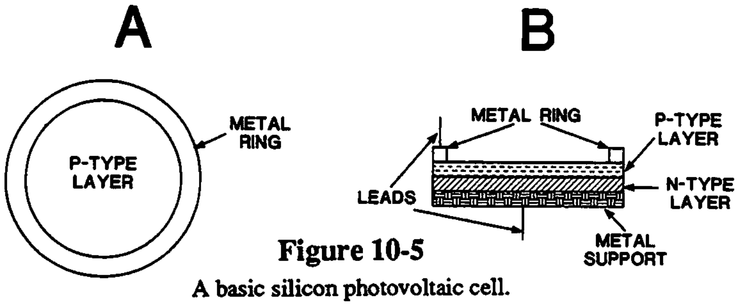


**Figure 10-5**
A basic silicon photovoltaic cell.
(A) topview (B) sideview

The photovoltaic cell is designed to have a relatively large surface area which can collect as much light as possible. The cell is constructed so that light must strike the top semiconductor layer within the metal ring as shown in the top view in Figure 10-5A. The device shown has a P-type upper layer but in some cases these cells are designed so that the N-type layer is on top and the P-type layer is on the bottom.

Since the photovoltaic cell has a PN junction, a depletion region (an area void of majority carriers) forms in the vicinity of the junction. If the cell was forward-biased like a conventional PN junction diode, the free electrons and holes in the device would be forced to combine at the junction and forward current would flow. However, the photovoltaic cell is not used in this manner. Instead of responding to an external voltage, the device actually generates a voltage in response to light energy which strikes its surface.

In order to generate a voltage, the top layer of the photovoltaic cell must be exposed to light. The light energy striking the cell is assumed to consist of many tiny particles or photons. These photons are actually absorbed at various depths into the semiconductor material depending on their wavelengths and energy content. If a particular photon has sufficient energy when it enters the semiconductor material, it can intercept an atom within the material and impart much of its energy to the atom. If sufficient energy is added to the atom, a valence electron may be knocked out of its orbit and become a free electron. This will leave the atom positively charged and a hole will be left behind at the valence site. In other words an impinging photon of adequate energy can produce an electron-hole pair. This corresponding electron and hole can drift through the semiconductor material. Additional electron-hole pairs are also produced by other photons which penetrate to various depths within the material.

Some of the free electrons and holes generated by the light energy are produced within the depletion region while others are generated outside of the region but are drawn into it. The free electrons in the region are swept from the P-type to the N-type material and the holes are drawn in the opposite direction. The electrons and holes flowing in this manner produce a small voltage across the PN junction, and if a load resistance is connected across the cell's leads, this internal voltage will cause a small current to flow through the load. This current will flow from the N-type material, through the load and back to the P-type material thus making the N and P regions act like the negative and positive terminals of a battery.

All of the photons striking the photovoltaic cell do not create electron-hole pairs and many of the electrons and holes which separate to form pairs eventually recombine. The cell is therefore a highly inefficient device as far as converting light energy into electrical power. When this efficiency is expressed in terms of electrical power output compared to the total power contained in the input light energy, most cells will have efficiencies that range from 3 percent up to a maximum of 15 percent.

As you might expect, the output voltage produced by a photovoltaic cell is quite low. These devices usually require high light levels in order to provide useful output power. Typical applications require an illumination of at least 500 to 1000 footcandles. At 2000 footcandles, the average open-circuit (no load connected) output voltage of a typical cell is approximately 0.45 volts. When loaded, a typical cell may provide as much as 50 or 60 milliamperes of output load current. However, by connecting a large number of cells in series or parallel, any desired voltage rating or current capability can be obtained.

When used on spacecraft or satellites many photovoltaic cells are connected together, as explained above, to obtain sufficient power to operate electronic equipment or charge batteries. However, these devices are also used as individual components in various types of test instruments and equipment. For example, they are used in portable photographic light meters (which do not require batteries for operation). They are also used in movie projectors to detect a light beam which is projected through the film. The light beam is modulated (controlled) by a pattern or sound track that is printed near the edge of the film. In this way, the intensity of the light beam is made to vary according to the sounds (voice and music) that occurred. The photovoltaic cell simply responds to the light fluctuations and produces a corresponding output voltage which can be further amplified and used to drive a loudspeaker which will convert the electrical energy back into sound. This application is shown in Figure 10-6.

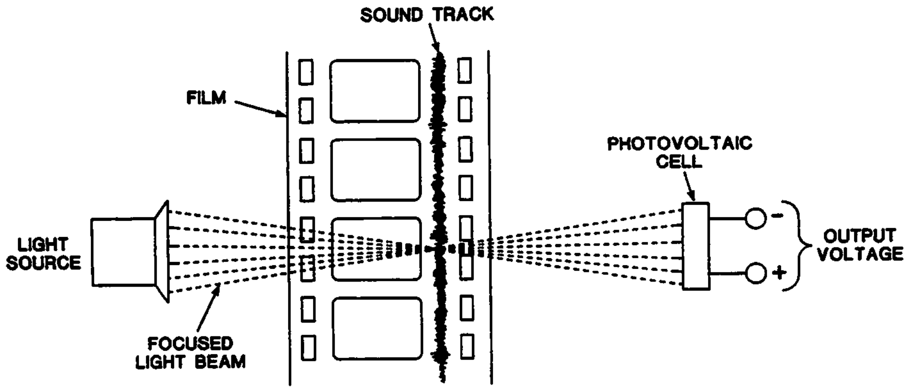


Figure 10-6
A photovoltaic cell used in a movie projector sound reproducing system.

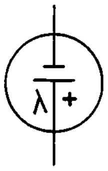


Figure 10-7
Commonly used symbol for a photovoltaic cell.

A schematic symbol that is commonly used to represent the photovoltaic cell is shown in Figure 10-7. This symbol indicates that the device is equivalent to a one-cell voltage source and the positive terminal of the device is identified by a plus (+) sign.

# Photodiodes

The photodiode is another light-sensitive device which utilizes a PN junction. It is constructed in a manner similar to the photovoltaic cell just described, but it is used in basically the same way as the photoconductive cell described earlier. In other words it is used essentially as a light-variable resistor.

The photodiode is a semiconductor device (usually made from silicon) and may be constructed in basically two ways. One type of photodiode utilizes a simple PN junction as shown in Figure 10-8A. A P-type region is diffused into an N-type substrate as shown. This diffusion takes place through a round window that is etched into a silicon dioxide layer that is formed on top of the N-type substrate. Then a metal ring or window is formed over the silicon dioxide layer (through an evaporation process) as shown. This window makes electrical contact with the P-type region and serves as an electrode to which an external lead can be attached. However, the window also accurately controls the area that will receive or respond to light. A metal base is then formed on the bottom N-type layer. This metal layer serves as a second electrode to which another lead is attached.

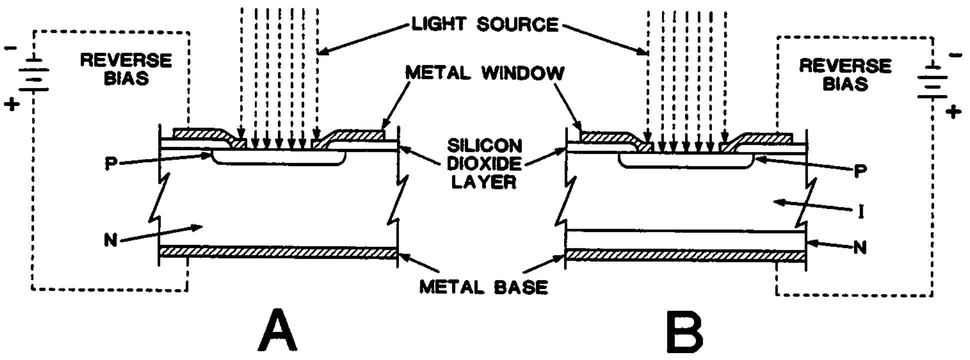


Figure 10-8
Basic construction of typical photodiodes.

The PN juction photodiode shown in Figure 10-8A operates on the same basic principle as the photovoltaic cell previously described. In fact, the photodiode may be used in basically the same manner as a photovoltaic cell. When used as a photovoltaic cell, the device is said to be operating in the *photovoltaic mode* and it will generate an output voltage (across its electrodes) that varies with the intensity of the light striking its top P-type layer. However, the photodiode is most commonly subjected to a reverse bias voltage as shown in Figure 10-8A. In other words its P-type region is made negative with respect to its N-type region. Under these conditions a wide depletion region forms around the PN junction. When photons enter this region to create electron-hole pairs, the separated electrons and holes are pulled in opposite directions because of the influence of the charges that exist on each side of the junction and the applied reverse bias. The electrons are drawn toward the positive side of the bias source (the N-type region) and the holes are attracted toward the negative side of the bias voltage (the P-type region). The separated electrons and holes therefore support a small current flow in the reverse direction through the photodiode. As the light intensity increases, more photons produce more electron-hole pairs which further increase the conductivity of the photodiode resulting in a proportionally higher current. When a photodiode is used in this manner it is said to be operating in the *photoconductive* or *photocurrent mode*.

The photodiode may also be constructed as shown in Figure 10-8B. This type of photodiode is similar to the type just described but there is one important difference. This device has an intrinsic (I) layer between its P and N regions and is commonly referred to as a PIN photodiode. The intrinsic layer has a very high resistance (a low conductivity) because it contains very few impurities. A depletion region will extend further into this I region than it would in a heavily doped semiconductor. The addition of the I layer results in a much wider depletion region for a given reverse bias voltage. This wider depletion area makes the PIN photodiode respond better to the lower light frequencies (longer wavelengths). The lower frequency photons have less energy content and tend to penetrate deeper into the diode's structure before producing electron-hole pairs and in many cases do not produce pairs. The wider depletion region in the PIN photodiode increases the chance that electron-hole pairs will be produced. The PIN photodiode is therefore more efficient over a wider range of light frequencies. The PIN device also has a lower internal capacitance due to the wide I region which acts like a wide dielectric between the P and N regions. This lower internal capacitance allows the device to respond faster to changes in light intensity. The wide depletion region also allows this device to provide a more linear change in reverse current for a given change in light intensity.

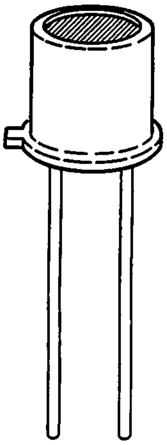


Figure 10-9A
A typical photodiode package.

PN junction and PIN photodiodes are often mounted on an insulative platform or substrate and sealed within a metal case as shown in Figure 10-9A. A glass window is provided at the top of the case, as shown, to allow light to enter and strike the photodiode. The two leads extend through the insulative base at the bottom of the case and are internally bonded (with fine wires) to the photodiode's electrodes.

The performance of a photodiode can be expressed in terms of its *quantum efficiency*. The quantum efficiency of a particular device is simply the ratio of the electrons produced for each photon that strikes the diode. Quantum efficiency can be expressed mathematically as:

$$\text{quantum efficiency} = \frac{\text{electrons}}{\text{photons}}$$

In the ideal case, one electron should be produced for each photon that strikes the diode thus giving the ideal diode a quantum efficiency of 1. However, the quantum efficiency of a typical photodiode will be lower than 1 and it will vary depending on the wavelength of the radiant energy.

Two photodiodes are mounted in the 10-lead metal can shown in Figure 10-9B, which has a glass window on top. With this package style, more leads are available than are actually needed.

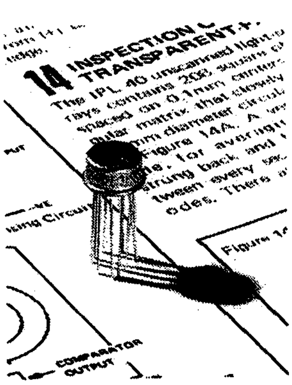


Figure 10-9B
A 10-lead metal can with two photodiodes. (courtesy of Integrated Photomatrix Inc.)

The quantum efficiency of a typical PIN photodiode is plotted graphically in Figure 10-10. The curve in this figure shows how quantum efficiency varies as wavelength varies. Notice that the quantum efficiency is approximately 0.3 (0.3 electrons per photon) at a wavelength of approximately 4000 angstroms and increases to a peak of approximately 0.8 at 8000 angstroms. Then it drops off to a value of 0.2 at 10,000 angstroms. This quantum efficiency curve effectively shows how well the photodiode responds to the various wavelengths. This curve shows that th diode's spectral response for all practical pruposes extends from approximately 4000 to 10,000 angstroms.

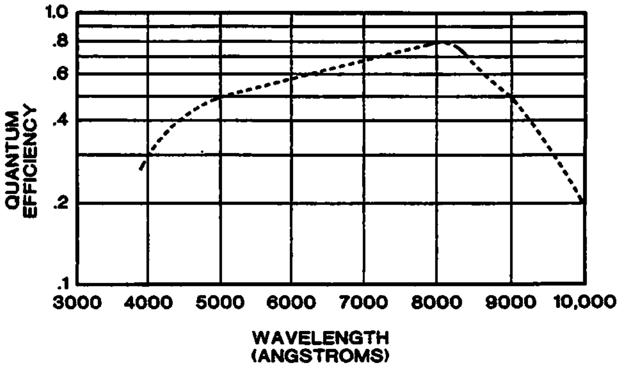


Figure 10-10
Spectral response
for a typical PIN photodiode.

The performance of a photodiode can also be expressed in terms of its *responsivity*. The responsivity of a photodiode is simply a measure of how much output (reverse) current is obtainable for a given light energy input. It is expressed as a ratio of output current (called *photocurrent*) in microamperes ($\mu$A) to the input radiant energy or irradiance which is measured in milliwatts per square centimeter ($mW/cm^2$). Expressed mathematically, responsivity is equal to:

$$\text{responsivity} = \frac{\mu A}{mW/cm^2}$$

The responsivity of a photodiode also varies with the wavelength of the radiant energy striking the device and reaches a peak value when the quantum efficiency of the device is near its highest value. A typical photodiode will have a maximum responsivity of 1.4 microamperes per milliwatt per square centimeter.

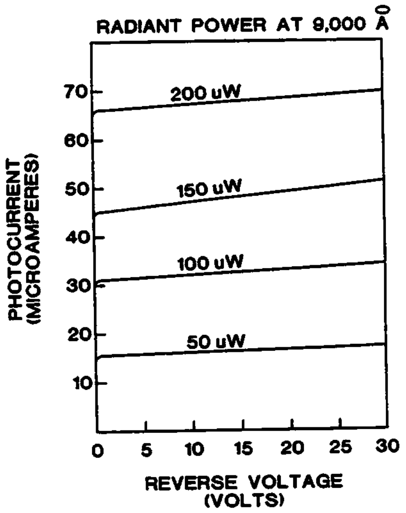


Figure 10-11
Typical photodiode electrical characteristics.

The electrical characteristics of a typical PIN photodiode are graphically shown in Figure 10-11. The curves in this figure show the relationship between the voltage across the photodiode and the current flowing through the device when it is used in the photoconductive mode which is the operating mode most widely used. The curves were plotted when the photodiode was exposed to radiant power at a wavelength of 9000 angstroms (0.9 microns) which is well within the diode's spectral response. Notice that the diode's photocurrent is plotted vertically while the reverse voltage applied to the diode is plotted horizontally. The lowest curve shows that when the diode is exposed to input radiant power of 50 microwatts (50 $\mu$W) its photocurrent remains almost constant (varies from approximately 15 to 18 microamperes) as its reverse voltage varies from 0 to 30 volts. The next higher curve shows that when the input light power is doubled (increased to 100 $\mu$W) the photocurrent almost doubles, but again remains essentially constant (slightly more than 30 microamperes) as reverse voltage varies from 0 to 30 volts. The two upper curves show that photocurrent continues to increase in proportion to the input light power (150 and 200 $\mu$W) even though the reverse voltage is varied. The curves in Figure 10-11 effectively show that the photodiode produces an output current that is relatively constant and this current is determined primarily by the input light power and to a lesser degree by the diode's reverse voltage. In most applications, the diode's reverse voltage is held constant and its output photocurrent is allowed to vary in proportion to the input light power thus allowing the device to operate in the photoconductive mode.

When the input radiant power is reduced to zero or when the photodiode is in complete darkness, the diode's photocurrent will drop to an extremely low value but not quite to zero. This very low current is simply a reverse leakage current which occurs in any type of PN junction device. This very low leakage current is referred to as the photodiode's *dark current*. The dark current is too small to be plotted in Figure 10-11 since it will generally be in the nanoampere range for most photodiodes. Dark current increases only slightly with an increase in reverse voltage.

Photodiodes have an important advantage over the photoconductive devices described earlier. A photodiode can respond much faster to changes in light intensity. In fact, the photodiode operates faster than any other type of photosensitive device. It is therefore useful in those applications where light fluctuates or changes intensity at a rapid rate. The major disadvantage with the photodiode is that its output photocurrent is relatively low when compared to other photoconductive devices.

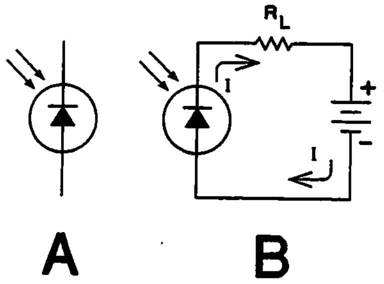


Figure 10-12
A photodiode symbol (A) and a properly biased photo diode (B).

Photodiodes and PIN photodiodes are both commonly represented by the same schematic symbol and several symbols have been used to represent these devices. A commonly used symbol is shown in Figure 10-12A. Notice that a conventional diode symbol is used with two arrows. The arrows point toward the diode to show that it responds to light. Figure 10-12B shows a properly biased photodiode. A load resistor ($R_L$) is also connected in series with the diode. The load resistor simply represents any resistive load which might be controlled by the photodiode as it varies its conductivity in accordance with input light intensity. The changes in the diode's conduction will cause the photocurrent (I) in the circuit to vary.

## Phototransistors

The phototransistor is also a PN junction device. However, it has two junctions instead of one like the photodiode just described. The phototransistor is constructed in a manner similar to an ordinary transistor, but this device is used in basically the same way as a photodiode.

The phototransistor is often constructed as shown in Figure 10-13. The process begins by taking an N-type substrate (usually silicon), which ultimately serves as the transistor's collector, and diffusing into this substrate a P-type region which serves as the base. Then an N-type region is diffused into the P-type region to form the emitter. The phototransistor therefore resembles a standard NPN bipolar transistor in appearance. The device is often packaged much like the photodiode shown in Figure 10-9. However, in the case of the phototransistor, three leads are generally provided which connect to the emitter, base, and collector regions of the device. Also, the phototransistor is physically mounted under a transparent window so that light can strike its upper surface as shown in Figure 10-13.

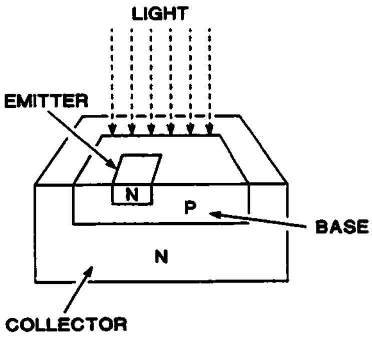


Figure 10-13
The construction of
a typical phototransistor.

The operation of a phototransistor is easier to understand if it is represented by the equivalent circuit shown in Figure 10-14. Notice that the circuit shown contains a photodiode which is connected across the base and collector of a conventional NPN bipolar transistor. If the equivalent circuit is biased by an external voltage source as shown, current will flow into the emitter lead of the circuit and out of the collector lead. The amount of current flowing through the circuit is controlled by the transistor in the equivalent circuit. This transistor conducts more or less depending on the conduction of the photodiode which in turn conducts more or less as the light striking it increases or decreases in intensity. If light intensity increases, the diode conducts more photocurrent (its resistance decreases) thus allowing more emitter-to-base current (commonly referred to as base current) to flow through the transistor. This increase in base current is relatively small but

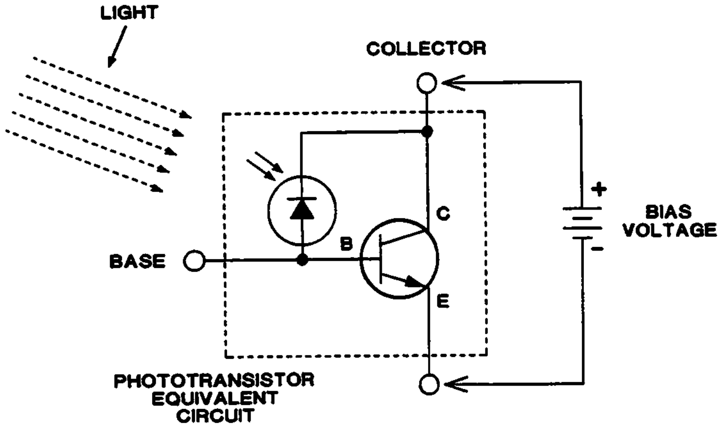


Figure 10-14
Equivalent circuit
for a phototransistor.

due to the transistor's amplifying ability this small base current is used to control the much larger emitter-to-collector current (also called collector current) flowing through this device. The increase in input light intensity causes a substantial increase in collector current. A decrease in light intensity would correspondingly cause a decrease in collector current.

Although the phototransistor has a base lead as well as emitter and collector leads, the base lead is used in very few applications. However, when the base is used, it is simply subjected to a bias voltage which will set the transistor's collector current to a specific value under a given set of conditions. In other words, the base may be used to adjust the phototransistor operating point. In most applications only the emitter and collector leads are used and the device is considered to have only two terminals.

The electrical characteristics of a typical phototransistor are graphically shown in Figure 10-15. The collector-to-emitter (collector) current is plotted vertically and the collector-to-emitter (collector) voltage is plotted horizontally. Each curve shown is plotted for a specific amount of irradiance or the amount of light striking the phototransistor per unit area. The lowest curve shows the relationship between collector current and collector voltage for an irradiance of 1 milliwatt per square centimeter. Notice that the collector current quickly jumps from 0 to just

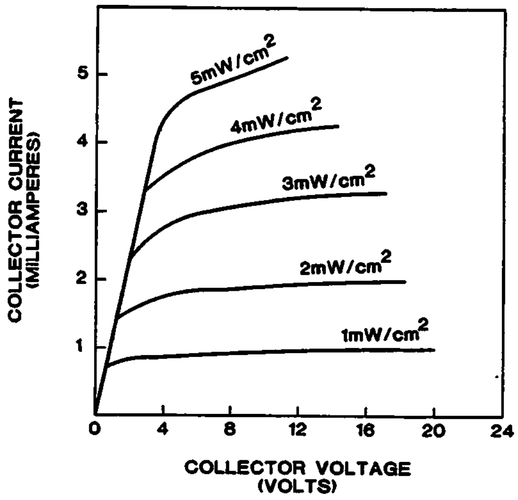


Figure 10-15
Typical phototransistor electrical characteristics.

slightly less than 1 milliampere and then increases to just slightly more than 1 milliampere as the collector voltage is varied from 0 to more than 16 volts. The collector current, although not constant, varies only a small amount over a wide range of collector voltages. The remaining curves are plotted for higher irradiance values and they effectively show that for each higher value the collector current is proportionally higher and increases only slightly as the collector voltage increases. The transistor's dark current is not shown since it is very low. As with the photodiode previously described, this current increases only slightly with an increase in operating voltage.

Figure 10-15 shows that the phototransistor, like the photodiode previously described, provides an output current that is essentially controlled by the intensity of the light striking its surface and to a lesser degree by its operating voltage. The phototransistor is used much like the photodiode by setting its collector voltage to a specific value and allowing the device to control its collector current in accordance with the changes in light intensity.

An important difference between the phototransistor and the photodiode is in the amount of current that each device can handle. The phototransistor can produce much higher output current than a photodiode for a given light intensity because the phototransistor has a built-in amplifying ability. The phototransistor's higher sensitivity makes it useful for a wider range of applications than a photodiode. Unfortunately this higher sensitivity is offset by one important disadvantage. The phototransistor does not respond as quickly to changes in light intensity and therefore is not suitable for applications where an extremely fast response is required. Like other types of photosensitive devices the phototransistor is used in conjunction with a light source to perform many useful functions. It can be used in place of photoconductive cells and photodiodes in many applications and can provide an improvement in operation. Phototransistors are widely used in such applications as tachometers, photographic exposure controls, smoke and flame detectors, object counting, and mechanical positioning and moving systems.

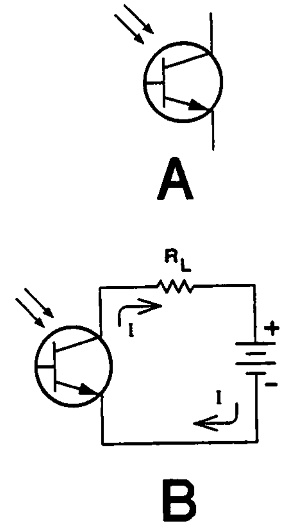


Figure 10-16
A phototransistor symbol (A) and a properly biased phototransistor (B).

A phototransistor is often represented by the symbol shown in Figure 10-16A and it is usually biased as shown in Figure 10-16B. As shown the phototransistor is used to control the current flowing through a load much like the photodiode shown in Figure 10-12B.

The phototransistor may also be used as a photodiode by using its collector and base leads and leaving its emitter open. When used in this manner, the PN junction between the collector and base region serves as a photodiode. However, when used as a photodiode, the device can handle only a relatively small current, although it will operate at a faster rate of speed.

A phototransistor may be interconnected with an ordinary bipolar transistor in a way which allows the phototransistor to control the operation of the bipolar transistor. This type of arrangement is referred to as a *photo-darlington* circuit and such a circuit can be simultaneously formed and packaged in a single container. The schematic symbol for a photo-darlington arrangement (in a single package) is shown in Figure 10-17 along with the necessary external circuitry. The phototransistor responds to the input light intensity and conducts accordingly. As it conducts more or less it controls the base current of the bipolar transistor which in turn controls the current through the load and the battery. Such an arrangement offers a tremendous increase in sensitivity since the gain of the phototransistor is effectively multiplied by the gain of the bipolar transistor thus producing a relatively high output current. However, the higher gain is obtained by sacrificing speed. Although it is more sensitive than the photodiode or phototransistor, it has a slower response to changes in light intensity.

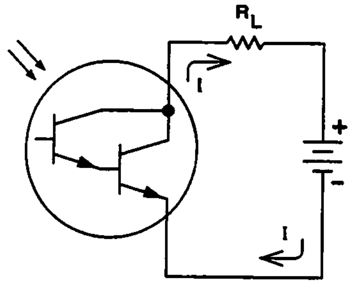


Figure 10-17
A properly biased photo-darlington circuit.

## Self Test Review

15. The photoconductive cell is basically a light-sensitive ____________.

16. The resistance of a photoconductive cell decreases as light intensity ____________.

17. The photovoltaic cell directly converts light into ____________ energy.

18. The photovoltaic cell is sometimes referred to as a ____________ cell.

19. Photovoltaic cells are often connected in ____________ to produce a higher output voltage.

20. The photodiode may operate in either the ____________ or ____________ mode.

21. A photodiode which has an intrinsic layer between its P and N regions is referred to as a ____________ photodiode.

22. The number of electrons produced for each photon which strikes a photodiode is referred to as the photodiode's ____________ ________.

23. The responsivity of a photodiode is a measure of how much reverse ____________ is obtainable for a given light energy input.

24. The term dark current refers to the amount of reverse leakage current flowing through a photo diode when the input radiant power is ____________.

25. The phototransistor provides a higher output ____________ than the photodiode.

26. The phototransistor's ____________ and ____________ leads are used in most applications.

27. When a phototransistor is connected to a bipolar transistor to obtain greater sensitivity the arrangement is referred to as a ____________ circuit.

## Answers

15. resistor
16. increases
17. electrical
18. solar
19. series
20. photovoltaic or photoconductive
21. PIN
22. quantum efficiency
23. current
24. zero
25. current
26. collector and emitter
27. photo-darlington

# EXPERIMENT 12

# Phototransistor Characteristics

*OBJECTIVE:* *To demonstrate the operation of a typical phototransistor.*

## Introduction

In this experiment, you will vary the amount of light applied to a phototransistor and observe how the resistance of the device changes with light intensity. Then, you will properly bias the phototransistor and demonstrate how the device controls the current flowing through a load resistance in accordance with light intensity. Finally, you will use the phototransistor as a photodiode and compare its operation with the phototransistor data obtained earlier.

## Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — Phototransistor FPT100 (417-851)
1 — Lamp (412-601)
1 — 10 kilohm, resistor (brown, black, orange, gold)

## Procedure

1. Locate your phototransistor and identify its emitter, base, and collector leads by referring to Figure 10-18. Spread the leads as shown so that they cannot touch each other and short out.

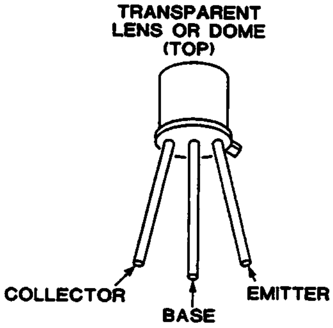


Figure 10-18

Identifying the leads of the phototransistor.

2. Connect your ohmmeter leads across the phototransistor's emitter and collector leads (negative lead on emitter) and hold the phototransistor so that your thumb covers the transparent dome or lens, so that no light can enter the device. Be sure to leave the base lead open. Measure the resistance of the phototransistor at this time and record your resistance value (in ohms) in the left column of the chart in Figure 10-19. Your recorded value represents the resistance of the phototransistor in complete darkness.

| PHOTOTRANSISTOR RESISTANCE | |
|---|---|
| NO LIGHT | NORMAL ROOM LIGHT |
| | |

**Figure 10-19**
Chart for recording
phototransistor resistance values.

3. Again measure the resistance of your phototransistor as you did in the previous step, but this time, do not cover the lens. This will allow the normal room light to enter the lens. Record your measured resistance value (in ohms) in the right column of the chart in Figure 10-19. Your recorded value represents the resistance of the phototransistor when it is exposed to normal room light.

## Discussion

In this first portion of the experiment, you simply measured the resistance of the phototransistor when it received no light and when it was exposed to normal room light. When the phototransistor received no light, its resistance should have been very high, since it could conduct very little current (only a small dark current) at that time. You may have obtained a resistance as high as 1 megohm, or possibly even higher. When you exposed the phototransistor to room light, its resistance should have dropped considerably, since more light makes the device conduct harder. Under these conditions, the resistance of the device may have been as low as 10 or 20 kilohms or possibly even lower. It is important to realize that the average light intensity in one room to the next can vary considerably and your phototransistor is very sensitive. It would, therefore be impossible to predict an exact, or even a close, resistance value. The important point to consider, is the tremendous drop in resistance as light intensity increases from zero to room level brightness. This simple test may be used to determine if a phototransistor is functioning properly.

## Procedure (continued)

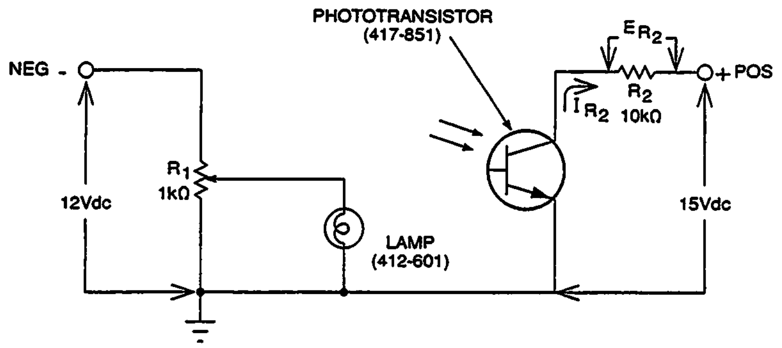


**Figure 10-20**

Circuit for experiment 12.

4. Construct the circuit shown in Figure 10-20 following the wiring diagram shown in Figure 10-21. Be sure to assemble the circuit as shown. Notice that the phototransistor is inserted into the breadboard socket and then pushed over on its side so that it is laying flush with the breadboard socket. The lamp must also be pushed down flat against the board so that the top of the lamp and the top (lens) of the phototransistor are approximately one-eighth of an inch apart. In this portion of the experiment, you will use the 1 kilohm potentiometer ($R_1$) to control the voltage applied to the lamp, and therefore, control its brightness. The phototransistor is positioned so that it will respond to the light produced by the lamp, and in turn, control the current through the 10 kilohm ($R_2$).

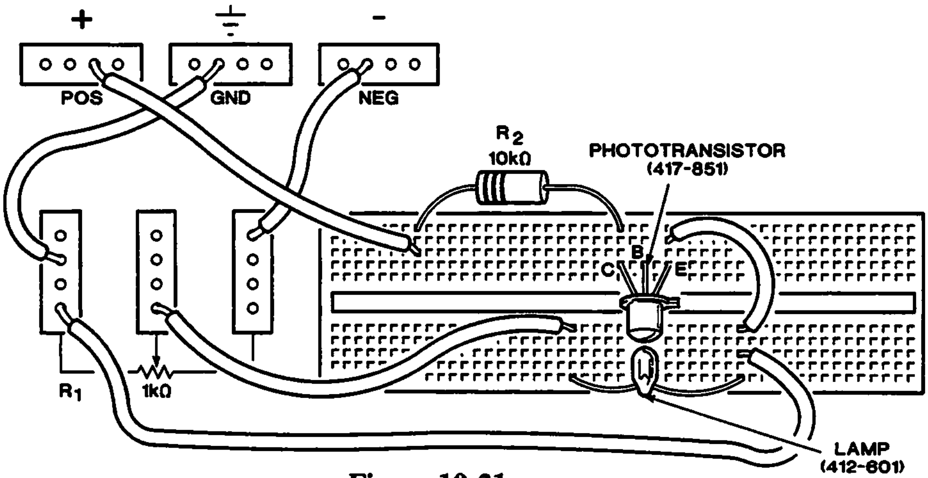


**Figure 10-21**

Wiring diagram for experimental circuit.

5. Adjust potentiometer $R_1$ for maximum resistance, and then, turn on your Trainer. Adjust the + voltage control for 15 volts. Then, adjust the – voltage control for 12 volts.

6. Measure the voltage across $R_2$. While taking this measurement, place your hand or some other object over the phototransistor to shield it from ambient room light. Try to reduce the light level to an absolute minimum. Record your measured voltage immediately below the designation $E_{R2}$ (left side) in the chart in Figure 10-22. Your recorded value represents the voltage across $R_2$ ($E_{R2}$) when the light intensity is essentially zero. Next use your recorded $E_{R2}$ value and the resistance of $R_2$ (10,000 ohms) and determine the current through $R_2$. Record your calculated current value in the $I_{R2}$ values are too low to accurately measure simply record a zero in the respective column in Figure 10-22.

| PHOTOTRANSISTOR CIRCUIT OUTPUT LOAD CURRENT AND VOLTAGE | | | |
|---|---|---|---|
| NO LIGHT | | MAX. LIGHT | |
| $E_{R2}$ | $I_{R2}$ | $E_{R2}$ | $I_{R2}$ |
| | | | |

**Figure 10-22**
Chart for recording output load current and voltage values for phototransistor circuit.

7. Now, unshield the phototransistor and measure the voltage across $R_2$ again. As you observe this voltage, slowly turn $R_1$ to minimum resistance (do not be concerned with normal room light). With $R_1$ at minimum resistance, measure the voltage across $R_2$ and record this voltage in the $E_{R2}$ column (right side) in Figure 10-22. Your recorded value represents the voltage across $R_2$ when the lamp intensity is highest. Calculate the current through $R_2$ and record your calculated value in the $I_{R2}$ column (right side) in Figure 10-22. Then, turn off your Trainer.

## Discussion

In this portion of the experiment, you demonstrated how a phototransistor can vary its conduction according to light intensity, and control the amount of current and voltage applied to a load resistance ($R_2$). First you shielded the phototransistor from light and determined the voltage across the load ($E_{R2}$) and the current through the load ($I_{R2}$). You found that the voltage across the load was very low at this time. It may have been only one or two tenths of a volt, or it could have been too low to accurately measure, in which case, it would be effectively zero. When you used this voltage value ($E_{R2}$) to determine $I_{R2}$, you should have obtained a very low current value. Depending on the value of $E_{R2}$, it may have been as low as 0.02 milliamperes (20 microamperes). If your phototransistor was adequately shielded, this current value will be approximately equal to the phototransistor's dark current value. An extremely low dark current is the mark of good phototransistor.

Next, you observed the voltage across the load as you increased the voltage applied to the lamp. As the lamp became brighter, you noticed that the voltage across the load ($E_{R2}$) increased in proportion to light intensity. When the light was fully on, you recorded the new value of $E_{R2}$ and calculated $I_{R2}$. Under these conditions $E_{R2}$ should have been equal to approximately 15 volts, indicating that the phototransistor was conducting fully and applying almost the full power supply voltage to the load. At this time, the load current ($I_{R2}$) should have been at its maximum possible value (approximately 1.5 milliamperes). This demonstration should clearly show that the phototransistor can be effectively used to control the current and voltage applied to a load in accordance with light intensity.

## Procedure (continued)

8. Remove the wire (one end only) that connects to the phototransistor's emitter lead, and reconnect this end of the wire to the phototransistor's base lead. Your phototransistor is now connected as a photodiode, since you are using only its collector and base leads.

9. Adjust $R_1$ to maximum resistance, and then turn on your Trainer.

10. Repeat Steps 6 and 7, but this time, record all of your measured and calculated values in the chart provided in Figure 10-23. When you are finished, turn off your Trainer.

| PHOTOTRANSISTOR CIRCUIT OUTPUT LOAD CURRENT AND VOLTAGE | | | |
|---|---|---|---|
| NO LIGHT | | MAX. LIGHT | |
| $E_{R2}$ | $I_{R2}$ | $E_{R2}$ | $I_{R2}$ |
| | | | |

**Figure 10-23**
Chart for recording output load current and voltage values for photodiode circuit.

## Discussion

In this final portion of the experiment, you used your phototransistor as a photodiode and observed the output current and voltage values. When the phototransistor was shielded, you probably found that the load voltage and current values ($E_{R2}$ and $I_{R2}$) were too low to measure, and were effectively zero. This proves that the dark current in a photodiode is extremely low. When you increased the intensity of the lamp to maximum, you should have noticed a slight increase in $E_{R2}$ and $I_{R2}$, however these values still remained quite low. You probably obtained an $E_{R2}$ value that was less than 1 volt, thus, indicating that $I_{R2}$ was less than 0.1 milliampere (100 microamperes). Your exact values are not important; however, your measurements and calculations should clearly show that the photodiode cannot produce an output current that is as high as that obtained with a phototransistor for a given input light intensity. The phototransistor, therefore, has a much higher responsivity (output current for a given light energy input) than a photodiode.

# Light-Emitting Devices

Light-emitting devices are components which produce light when they are subjected to an electrical current or voltage. In other words, they simply convert electrical energy into light energy. For many years the incandescent lamp and the neon lamp were the most popular sources of light in various electrical applications. The incandescent lamp simply uses a metal filament which is placed inside of a glass bulb and air is drawn out of the bulb to produce a vacuum. Current flows through the filament causing it to heat up and produce light. The neon lamp utilizes two electrodes which are placed within a neon gas filled bulb. In this case, current is made to flow from one electrode to the other through the gas, the gas ionizes, and emits light.

The incandescent lamp produces a considerable amount of light, but its life expectancy is quite short. A typical lamp might last as long as 5000 hours. In addition to having a short life, the incandescent lamp responds slowly to changes in input electrical power. The incandescent lamp was (and still is) suitable for use as an indicator, or for simply providing illumination; but, due to its slow response, it will not faithfully vary its light intensity in accordance with rapidly changing alternating currents. The incandescent lamp, therefore, cannot be effectively used to convert high frequency electrical signals (much above the audio range) into light energy which is suitable for transmission through space. The light energy produced by the incandescent lamp is not useful for carrying information which could subsequently be recovered or converted back into an electrical signal by a suitable light sensitive device.

The neon lamp has a somewhat longer life expectancy (typically 10,000 hours) than an incandescent lamp and a somewhat faster response to changes in input current. However, its output light intensity is much lower than that of an incandescent device. The neon lamp has been used for many years as an indicator or warning light, and in certain applications, to transmit low frequency ac signals or information in the form of light over very short distances.

With all of their shortcomings, incandescent and neon lamps were used for many years simply because nothing better was available. However when the light-emitting diode (LED) was developed, it revolutionized the optoelectronics field. This device is a solid-state component, and it is physically stronger than the glass encased incandescent and neon lamps. Like all semiconductor devices, it has an unlimited life expectancy. Since it is such an important solid-state component, we will examine the operation and construction of an LED in detail. Then, we will see how it is used in various applications.

# LED Operation

You have learned how light energy (photons) striking a PN junction diode can impart enough energy to the atoms within the device to produce electrons and holes are swept across the diode's junction and support a small current through the device. However, the exact opposite is also possible. A PN junction diode can also emit light in response to an electric current. In this case, light energy (photons) is produced because electrons and holes are forced to recombine. When an electron and hole recombine, energy may be released in the form of a photon. The frequency (or wavelength) of the photons emitted in this manner is determined by the type of semiconductor material used in the construction of the diode.

The LED utilizes the principle just described. It is simply a PN junction diode that emits light through the recombination of electrons and holes when current is forced through its junction. The manner in which this occurs is illustrated in Figure 10-24. As shown in this figure, the LED must be forward-biased so that the negative terminal of the battery will inject electrons into the N-type layer (the cathode and these electrons will move toward the junction. Corresponding holes will appear at the P-type or anode end of the diode (actually caused by the movement of electrons) and also appear to move toward the junction. The electrons and holes merge toward the junction where they may combine. If an electron possesses sufficient energy when it fills a hole, it can produce a photon of light energy. Many such combinations can result in a substantial amount of light (many photons) being radiated from the device in various directions.

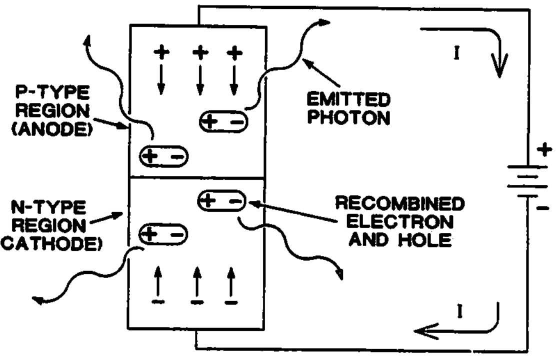


**Figure 10-24**
Basic Operation of a light emitting diode.

At this time, you are probably wondering why the LED emits light and an ordinary diode does not. This is simply because most ordinary diodes are made from silicon and silicon is an opaque or impenetrable material as far as light energy is concerned. Any photons that are produced in an ordinary diode simply cannot escape. LED's are made from semiconductor materials that are semitransparent to light energy. Therefore, in an LED some of light energy produced can escape from the device.

## LED Construction

Many LED's are made of gallium arsenide (GaAs). The LED's made from this material emit light most efficiently at a wavelength of approximately 9000 angstroms which is in the infrared region of the light spectrum and is not visible to the human eye. Other materials are also used such as gallium-arsenide phosphide (GaAsP) which emits a visible red light at approximately 6600 angstroms and gallium phosphide (GaP) which produces a visible green light at approximately 5600 angstroms. The GaAsP device also offers a relatively wide range of possible output wavelengths by adjusting the amount of phosphide in the device. By adjusting the percentage of phosphide, the LED can be made to emit light at any wavelength between approximately 5500 angstroms to 9100 angstroms.

Although Figure 10-24 helps to illustrate the operation of an LED, it does not show how the device is constructed. The construction of a typical GaAsP LED is shown in Figure 10-25. Figure 10-25A shows a cross-section of the device and Figure 10-25B shows the entire LED chip. The construction begins with a gallium arsenide (GaAs) substrate. On this substrate an epitaxial layer of gallium arsenide phosphide (GaAsP) is grown, however the concentration of gallium phosphide GaP) in this layer is gradually increased from zero to the desired level. A gradual increase is required so that the crystalline structure of the substrate is not disturbed. During this growth period, an N-type impurity is added to make the epitaxial layer an N-type material. The grown layer is then coated with a special insulative material and a window is etched into this insulator. A P-type impurity is then diffused through the window into the epitaxial layer and the PN junction is formed. The P-type layer is made very thin so that the photons generated at or near the PN junction will have only a short distance to travel through the P-type layer and escape as shown in Figure 10-25A.

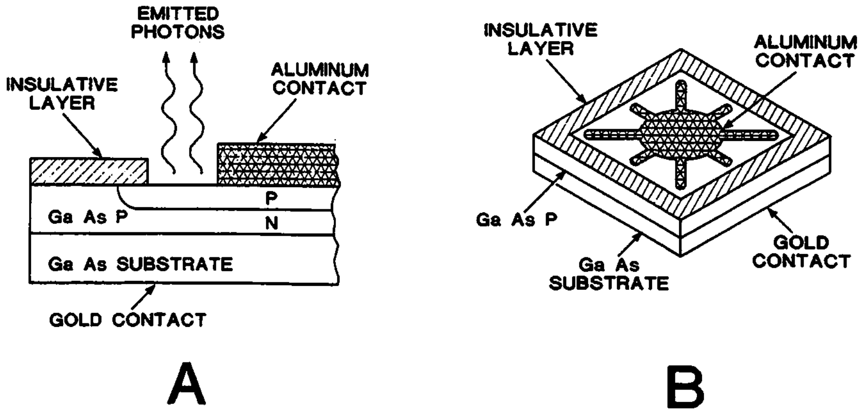


Figure 10-25

Basic construction of an LED.

The construction of the GaAsP LED is completed by attaching electrical contacts to the P-type region and the bottom of the substrate. The upper contact has a number of fingers extending outward so that current will be distributed evenly through the device when a forward bias voltage is applied across the contacts.

Once the LED is formed it must be mounted in a suitable package. Several types of packages are commonly used but all must fulfill one important requirement. All packages must be designed to optimize the emission of light from the LED. This factor is very important because the LED emits only a small amount of light. Therefore most packages contain a lens system which gathers and effectively magnifies the light produced by the LED. Also, various package shapes are used to obtain variations in the width of the emitted light beam or variations in the permissible viewing angle.

A typical LED package is shown in Figure 10-26. As shown, the package body and lens are one piece and are molded from plastic. The cathode and anode leads are inserted through the plastic case and extend up into the dome shaped top which serves as the lens. The bottom contact of the LED chip is attached directly to the cathode lead and the upper contact is connected to the anode lead by a thin wire which is bonded in place. The placement of the LED chip in the case is critical since the case serves as a magnifier. In some cases the plastic lens will contain fine particles which help to diffuse the light or the entire case may be dyed or tinted with a color that enhances the natural light color emitted by the LED.

The LED package shown in Figure 10-26 is installed by simply pushing the lens through a suitable hole in a chassis or special bezel and snapping it in place. The leads are then soldered into place. A notch or flat edge indicates the cathode lead. Also in most cases, the cathode lead will be shorter than the anode lead.

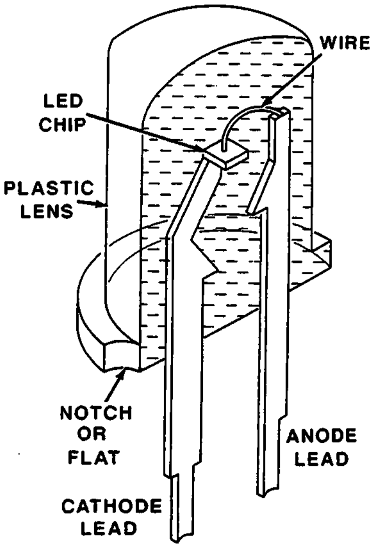
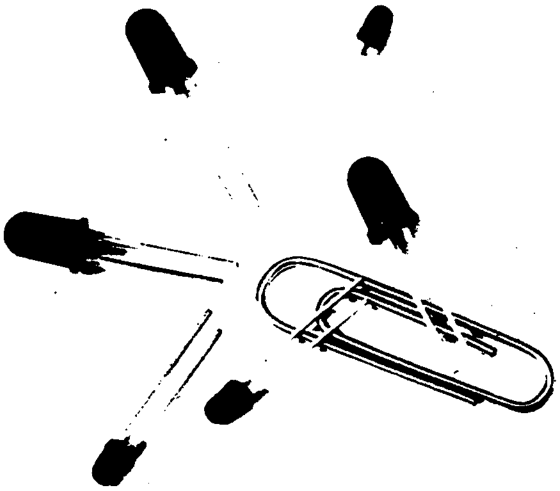


Figure 10-26
A typical LED package.



Typical light-emitting diodes (LED's). Most LED's produce red light, but special devices to generate green or yellow lights are also available. (Courtesy of Hewlett-Packard)

## LED Characteristics

The relationship between forward current and voltage in a typical GaAsP LED is shown graphically in Figure 10-27. Notice that the forward bias must be increased to approximately 1.2 volts before any appreciable forward current flows. Then, current increases rapidly for a continued increase in forward voltage. This graph effectively shows that once the LED conducts, its current can vary over a considerable range up to its maximum value while the voltage across the LED remains essentially constant at approximately 1.6 volts. Most LED's exhibit a similar current-voltage relationship.

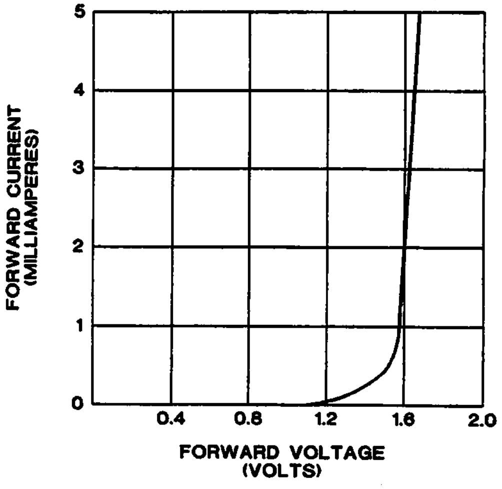


Figure 10-27
Current-Voltage Characteristics of a typical GaAsP LED.

Figure 10-28 shows the relationship between the foward current and the total output radiant power produced by a typical GaAsP LED. The power is expressed only in relative terms (as a percentage). The important point to note in Figure 10-28 is that the output radiant power increases linearly with the forward current.

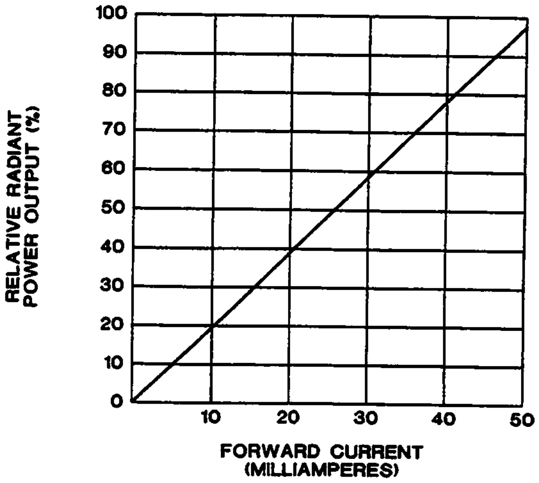


Figure 10-28
Output radiant power versus forward current for a typical GaAsP LED.

The spectral response of a typical GaAsP LED is shown in Figure 10-29. This graph shows the relationship between output radiant power and the wavelength of the radiated energy. The output power is only relative and is plotted as a percentage of the maximum possible. Notice that the GaAsP LED provides the highest output at approximately 6600 angstroms and the output rapidly drops off on either side of this maximum point. This relatively narrow spectral output produces a red light as explained earlier.

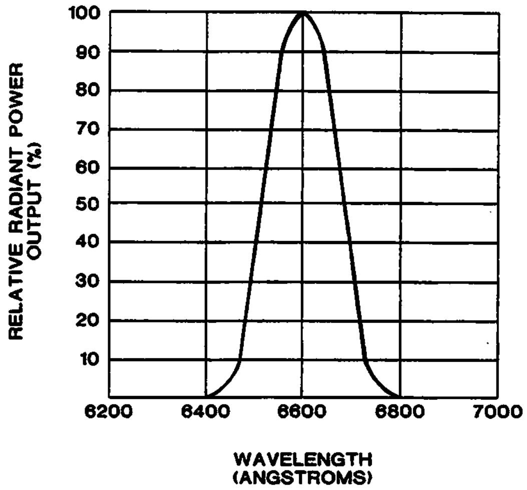


Figure 10-29
Spectral response of a typical GaAsP LED.

The performance of an LED can be expressed in terms of its efficiency or the number of output photons produced for each input electron. When its efficiency is expressed in this manner, it is found to be extremely low for all types of LED's. This low efficiency occurs for two basic reasons. First, much of the light energy produced never reaches the surface of the device where it can escape. It is simply absorbed by the semitransparent semiconductor material. The longer the path through which the light must travel, the greater will be the absorption. This is why the P-type region is made very thin. The thin P-type layer places the junction near the surface where light can escape. Unfortunately the emitted photons tend to move out in all directions from the junction. However, this situation is corrected by the gold contact or backing (refer to Figure 10-25) on the bottom of the chip. This backing reflects the light emitted toward the bottom of the chip back towards the P-type layer and helps to increase the efficiency of the device.

The low efficiency of an LED also results because much of the light is reflected back into the device. In order for the photons to escape, they must strike the surface within a specified *critical angle* as shown in Figure 10-30. For the GaAsP LED the critical angle is 17°. This simply means that photons must strike the surface within a 17° angle (from the line drawn perpendicular to the surface) or be reflected back into the structure. The critical angle of 17° is true only when the photons are emitted into the air and under those conditions only 8 percent of the light produced can escape.

When the LED chip is placed in a plastic (epoxy is sometimes used) package the critical angle is greatly increased and in some cases can be doubled. The package therefore increases the amount of light emitted in addition to magnifying or diffusing the light so that it can be easily seen.

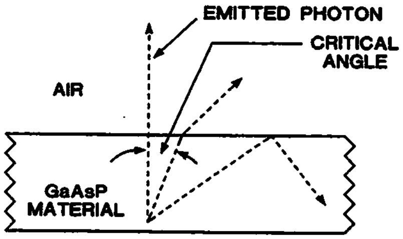


Figure 10-30

Relative paths of photons arriving at the surface of an LED chip.

As you may suspect, the amount of light produced by an LED is small compared to an incandescent lamp. Most LED's generate a typical luminous intensity of only a few millicandelas, which is very low compared to even a miniature incandescent panel light which can produce many times that much light. However, LED's have several important advantages. First, they are extremely rugged. They also respond very quickly to changes in operating current and therefore can operate at extremely high speeds. They require very low operating voltages and are therefore compatible with integrated circuits, transistors, and other solid-state devices. They are relatively inexpensive when compared to incandescent devices. Also, they may be designed to emit a specific light color or narrow frequency range as compared to the incandescent lamp which emits a white light that contains a broad range of light frequencies.

The disadvantages associated with LED's (in addition to low light output) are similar to those which pertain to many types of solid-state components. They may be easily damaged by excessive voltage or current (beyond their maximum ratings) and their output radiant power is dependent on temperature.

# LED Applications

In any application of an LED, the device is seldom used alone. The LED is usually connected in series with a resistor which limits the current flowing through the LED to the desired value. To operate the LED without this current limiting resistor would be risky since even a slight increase in operating voltage might cause an excessive amount of current to flow through the device. Some LED packages even contain built-in resistors (in chip form).

A schematic symbol that is commonly used to represent the LED is shown in Figure 10-31A and the correct way to bias an LED is shown in Figure 10-31B. The series resistor ($R_S$) must have a value which will limit the forward current ($I_F$) to the desired value based on the applied voltage (E) and the voltage drops across the LED which we may assume to be equal to 1.6 volts. When determining the value of $R_S$ allowance must also be made for the internal resistance of the LED (which has a typical value of 5 ohms). All of these factors are included in the following equation which can be used to determine the required value of $R_S$:

$$R_S = \frac{E - 1.6}{I_F} - 5\ \Omega$$

If we assume that E is equal to 6 volts and the $I_F$ must equal 50 milliamperes (0.05 amperes) to obtain the desired light intensity, $R_S$ must be equal to:

$$R_S = \frac{6 - 1.6}{0.05} - 5\ \Omega$$

$$R_S \doteq 88 - 5$$

$$R_s = 83 \text{ ohms}$$

The previous equation is useful for determining $R_S$ as long as the required $R_S$ value is equal to or greater than 40 ohms.

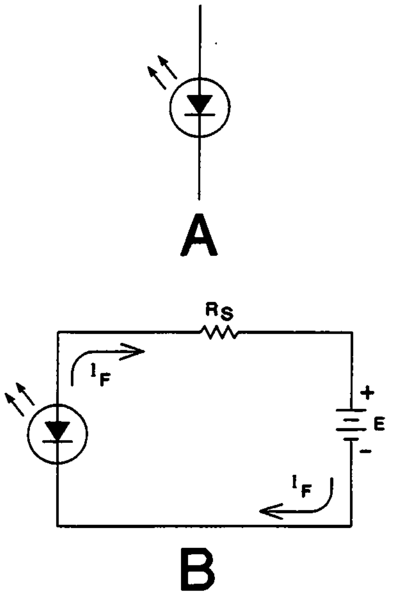


Figure 10-31
An LED symbol (A)
and a properly biased LED circuit (B).

Visible light producing LED's with their respective series resistors are often used as simple indicator lights to provide simple on and off indications. Individual LED's may even be arranged into specific patterns as shown in Figure 10-32. The LED's shown in this figure each illuminate one of seven segments arranged in a special pattern. The segments can be turned on or off to create the numbers 0 through 9 and certain letters. Such devices are referred to as 7 segment LED *displays*.

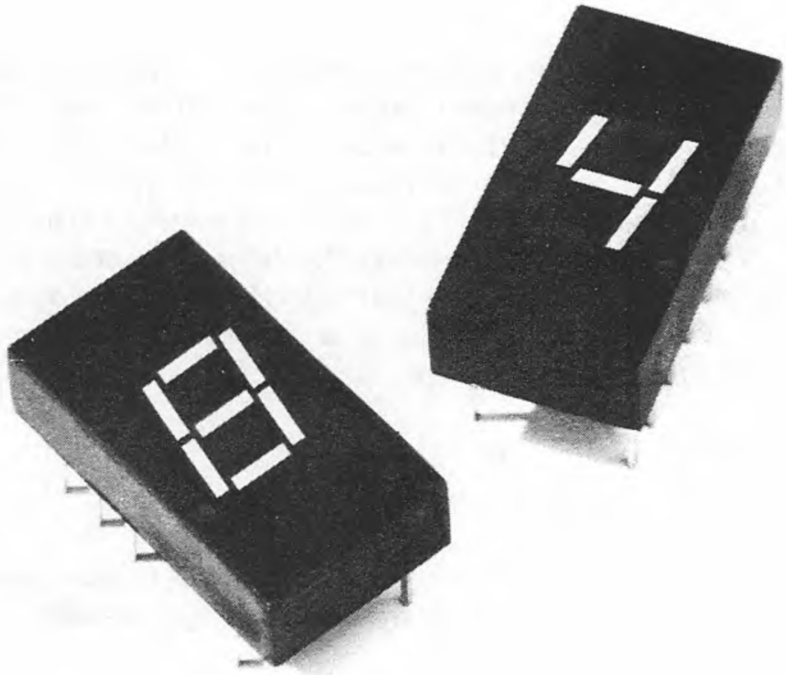


Figure 10-32
Typical segment LED numeric displays in a standard dual in line package (Courtesy of Hewlett-Packard)

LED's that emit infrared light may be used in intrusion detector systems if the light is properly focused and controlled. The infrared light cannot be seen by the human eye and is very effective in this application.

Infrared LED's are also commonly used in conjunction with light-sensitive devices, such as photodiodes or phototransistors, to form what is called an *optical coupler*. A typical optical coupler, which utilizes an LED and a phototransistor, is shown in Figure 10-33. The LED and phototransistor chips are separated by a special type of light transmitting glass, and they are coupled only by the light beam produced by the LED. An electrical signal (varying current or voltage) applied to the LED's terminals (through two of the pins on the mini-DIP package) will produce changes in the light beam which in turn varies the conductivity of the phototransistor. When properly biased, the phototransistor will convert the varying light energy back into an electrical signal. This type of arrangement allows a signal to pass from one circuit to another but provides a high degree of electrical isolation between the circuits. Also, the LED responds quickly to input signal changes thus making it possible to transmit high frequency ac signals through the optical coupler.

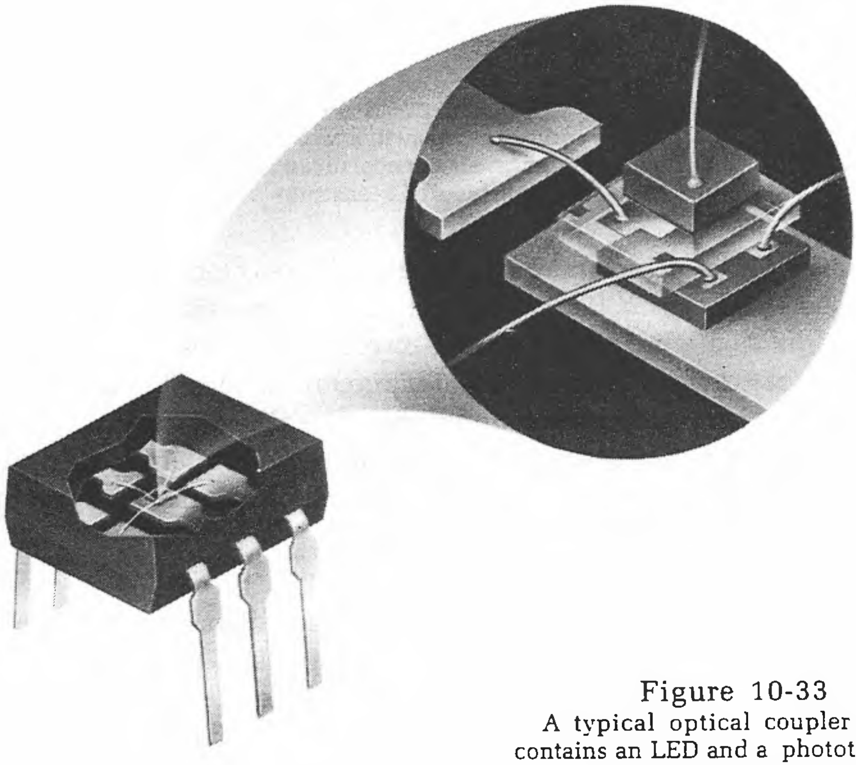


Figure 10-33
A typical optical coupler which contains an LED and a phototransistor. (Courtesty of General Electric)

Another application of the LED is shown in Figure 10-34. The device shown in this figure is called an *optical limit switch*. An LED is mounted in the right portion of the package and a photo-darlington circuit is mounted on the left side. They are separated by a narrow slot in the package. The device may be thought of as an optical coupler with an exposed light beam. When an object (such as a dime, as shown) is inserted into the slot, the light beam is broken and the photo-darlington circuit does not provide an output signal. This condition is identified by an LED indicator lamp which is mounted above the package. When the object is removed the light beam is restored and the indicator lamp turns on as shown in the right photo. Such a device is useful for sensing the presence of objects such as cards, tickets or tapes.

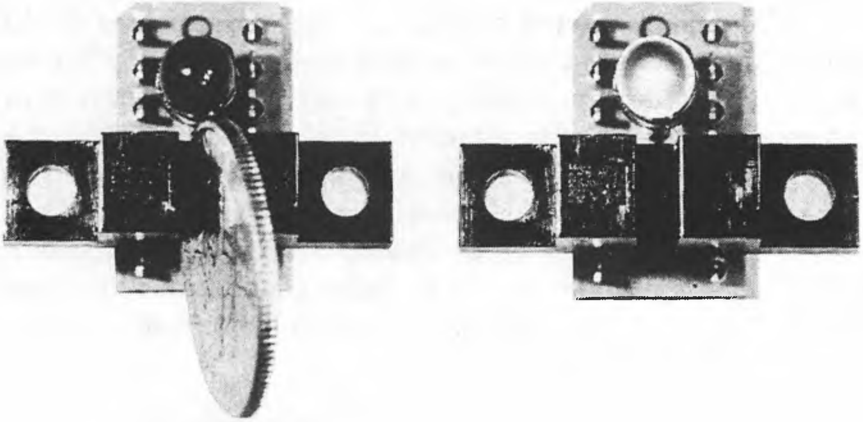


Figure 10-34
An optical limit switch which senses the presence or absence of an object.
(Courtesy of Monsanto)

LED's are also used in conjunction with photodiodes or phototransistors to sense the presence of holes or perforations in paper tapes or cards. However, in this application a number of LED's as well as photodiodes or phototransistors are required and they are arranged in specific patterns to correspond with the holes in the tape or card.

# LIQUID CRYSTAL DISPLAYS

Liquid-crystal displays are considered light transmissive or light reflective devices. Liquid-crystal displays are usually packaged as numeric or alphanumeric displays similar to those often used to package LED's. However, the liquid-crystal display (LCD) differs from an LED display in one important respect, LED's generate light while LCD's control light. Consequently, the appearance of an LCD is quite different from that of an LED. LCD characters are usually dark against a light background. A significant advantage of LCD displays is that they have very low operating voltage and current requirements. As a result, these displays have become very popular in digital wrist watches, test equipment, calculators, and other battery powered devices. Figure 10-35 shows a portable multimeter that uses an LCD readout.

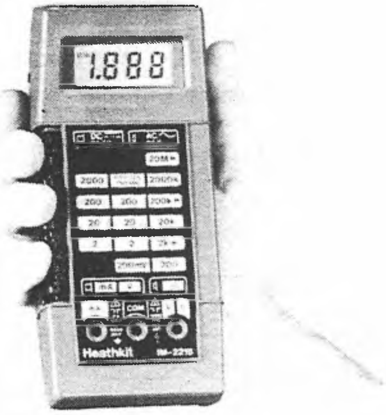
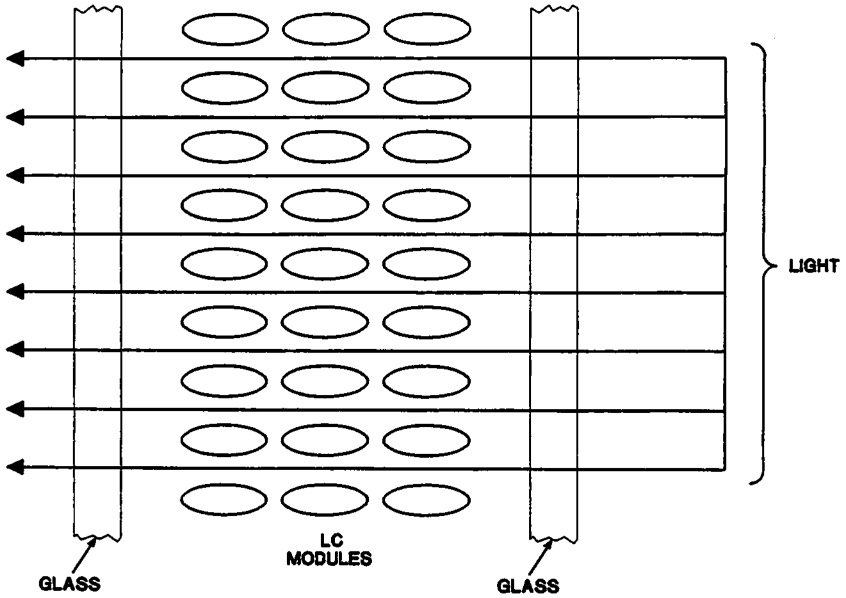
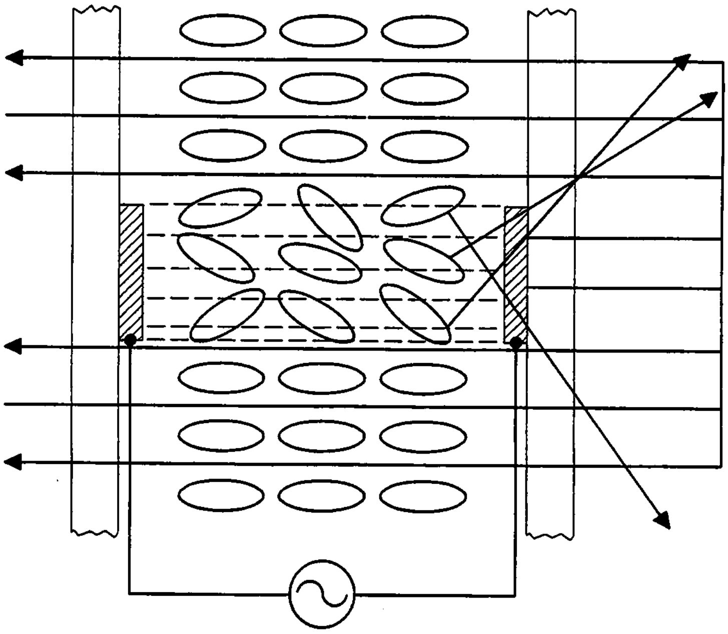


**Figure 10-35**
Typical hand-held multimeter,
using an LCD readout.

The term "liquid crystal" may seem to be contradictory because a crystalline structure is usually associated with solids. However, there is an unusual group of fluids that has a molecular structure like that of crystalline solids. The fluids in this group that are used in optoelectronic displays are *nematic* liquid crystals. Nematic liquid crystals have long molecules that are ordinarily aligned in a specific pattern. As shown in Figure 10-36A, light can readily pass through this molecular structure. Therefore, the material is transparent. However, as indicated in Figure 10-36B, if an electrical or magnetic field is applied to all or part of the liquid crystal, the molecular arrangement in that part of the crystal is subjected to the field changes. This molecular rearrangement causes scattered reflection of light with the result that the material takes on an opaque, frosty appearance.



**A** NO FIELD APPLIED



**B** WITH FIELD APPLIED

10-36
Dynamic scattering effect.

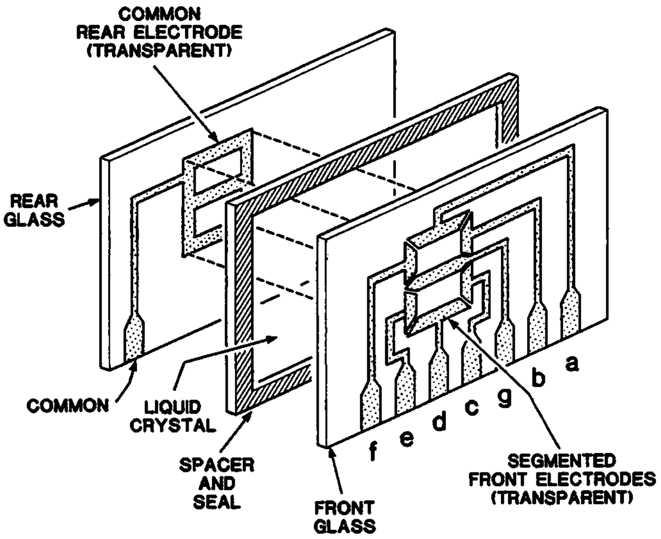


Figure 10-37
Exploded view of LCD showing construction details.

Figure 10-37 shows the construction of a typical liquid-crystal display. As shown, the liquid-crystal material is sealed between two layers of glass. Transparent electrodes have been deposited on each layer of glass. In addition, the molecules in the nematic liquid crystal are aligned in a uniform lattice arrangement that is perpendicular to the glass, allowing light to pass through the display.

Application of a voltage to one or more front electrodes and the rear electrode produces an electrostatic field, forcing the liquid crystal molecules within the electrostatic field to rearrange themselves into a random alignment pattern. As a result, light striking this part of the display is scattered, causing it to appear opaque. With the addition of a dark, light absorbing background to the display, frosty white characters will stand out clearly against a dark background.

Although LCD's have many advantages over other types of displays, LCD's have the disadvantage that they must be illuminated by an external light source. Consequently, they cannot be read in the dark or in low light levels. In some digital wrist watches, and other similar devices, this disadvantage is partially overcome by installing a small light source in or near the display. The user can switch the light source on and off as required to read the display.
)

# Self Test Review

28. Light-emitting devices convert electrical energy into ________________ energy.

29. Incandescent lamps have a very long life expectancy.

    A. True
    B. False

30. The light-emitting diode must be forward biased to operate properly.

    A. True
    B. False

31. A properly biased LED emits light because of the recombination of __________ and __________ near its PN junction.

32. LED's made with gallium arsenide emit light in the ________ region.

33. Gallium phosphide LED's produce a visible ____________ light.

34. The construction of the LED is completed by attaching the electrical contacts. The upper contact has a number of ______ extending outward, so that current will be distributed evenly through the device.

35. The LED package usually has a dome shaped top which serves as a _________.

36. The radiant power produced by an LED increases linearly with its input ____________.

37. The low efficiency of an LED results because much of the light generated within the device is either absorbed or ___________ back into its structure.

38. In order for the photons to escape from the surface of an LED they must strike the surface within a specified __________ ___________.

39. A _________ is normally placed in series with an LED to limit its current to the desired value.

40. LED's that are arranged into patterns so that they will create numbers are referred to as LED ___________.

41. An optical coupler containing an LED and a photodiode will transmit information over a light beam but at the same time provide a high degree of electrical __________.

42. In the construction of an LCD, the liquid-crystal material is sealed between two layers of glass. Transparent ___________ are deposited on each layer of glass.

43. A significant advantage of LCD displays are their ___________ operating voltages and ____________ operating current requirements.

## Answers

28. light
29. b. False
30. a. True
31. electrons and holes
32. infrared
33. green
34. fingers
35. lens
36. current
37. reflected
38. critical angle
39. resistor
40. displays
41. isolation
42. Electrodes
43. Low; Low.

# EXPERIMENT 13

# LED Characteristics

OBJECTIVE: *To demonstrate the operation of a typical visible light-emitting diode.*

## Introduction

In this experiment, you will observe the operation of a typical GaAsP LED which emits a visible red light. You will observe the relationship between the LED's forward current and forward voltage as you observe the LED's brightness. You will also reverse bias the LED and observe its operation.

## Material Required

Heathkit Analog Trainer
VOM (Reference lead must not be earth grounded.)
1 — LED (412-640)
1 — 1 kilohm resistor (brown, black, red, gold)

## Procedure

1. Construct the circuit shown in Figure 10-38 following the wiring diagram shown in Figure 10-39. Examine Figure 10-39 closely so that you will be sure to install the LED properly. Notice that the LED has a flat side (or possibly a notch) which identifies its cathode lead.

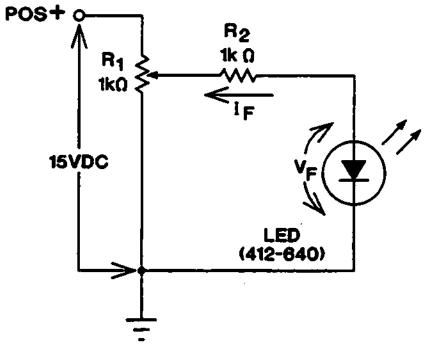


**Figure 10-38**

Circuit for experiment.

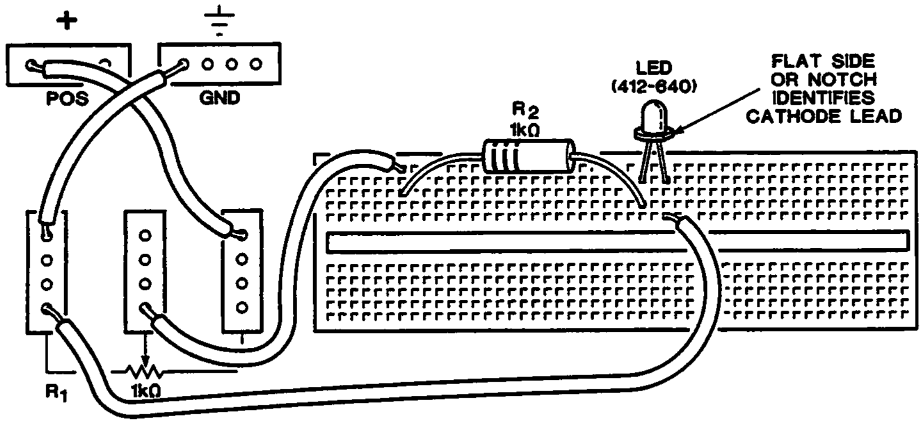


**Figure 10-39**

Wiring diagram for experimental circuit.

2. Adjust potentiometer $R_1$ for maximum resistance. Potentiometer $R_1$ will be used to control the voltage applied to the LED and its series 1 kilohm resistor ($R_2$). As you continue this experiment, you will increase this input voltage and observe the voltage across the LED and the current flowing through it. As you perform the following steps, be sure to notice the variations in the LED's brightness.

3. Turn on your Trainer and adjust the + voltage control for +15 volts across $R_1$.

4. Now turn $R_1$ towards minimum resistance until the voltage across $R_2$ is equal to 1 volt. Since $R_2$ has a value of 1 kilohm, the current through it will equal:

$$I = \frac{1V}{1000\,\Omega} = 1 \text{ milliampere}$$

   Since the same current is flowing through the LED, the LED's forward current ($I_F$) must now be equal to 1 milliampere.

5. Measure the forward voltage ($V_F$) across the LED. Record your measured value (in volts) immediately below the $I_F$ value of 1 milliampere in the table provided in Figure 10-37. Your recorded value indicates the amount of forward voltage ($V_F$) across the LED when its forward current ($I_F$) is equal to 1 milliampere.

6. Complete the table in Figure 10-40 by determining the $V_F$ value which corresponds to each $I_F$ value given. To do this, simply repeat steps 4 and 5 for each $I_F$ value given. Remember to adjust the voltage across $R_2$ to 2, 3, or 4 volts, etc., to obtain $I_F$ values of 2, 3, or 4 milliamperes, respectively. When you finish, turn off the Trainer.

| LED CURRENT-VOLTAGE MEASUREMENTS | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $I_F$ (MILLIAMPERES) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $V_F$ (VOLTS) | | | | | | | | | | |

**Figure 10-40**
Chart for recording $I_F$ and $V_F$ values.

7. Now plot the corresponding $I_F$ and $V_F$ values in Figure 10-40 on the graph shown in Figure 10-41. Connect the points to form a continuous curve.

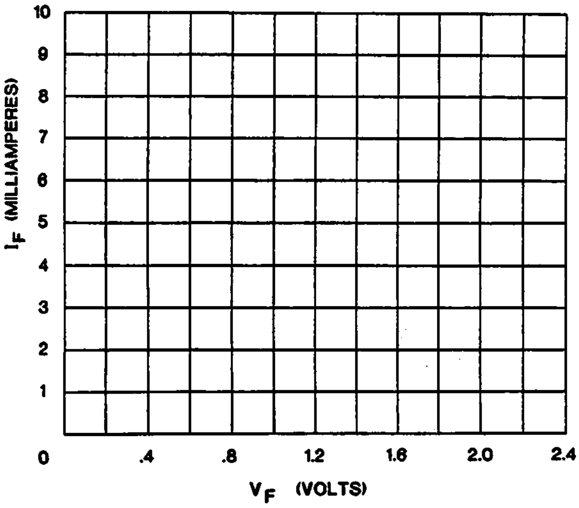


**Figure 10-41**

Graph for plotting $I_F$ and $V_F$ values.

## Discussion

In this part of the experiment, you gradually increased the forward current ($I_F$) through the LED and determined its forward voltage ($V_F$) at intermediate points. Then you plotted a curve which effectively shows the relationship between $I_F$ and $V_F$ remains relatively constant but does increase slightly as $I_F$ increases. Your $V_F$ values probably started at some value between 1.5 and 2 volts and increased to a few tenths of a volt as $I_F$ increased to its maximum value. The exact $V_F$ values are not important, but your curve should have the same general shape as the one shown in Figure 10-27. It should rise rapidly showing that $V_F$ values are not important, but your curve should have the same general shape as the one shown in Figure 10-27. It should rise rapidly showing that $V_F$ is essentially constant with changes in $I_F$. The lower portion of your curve (between 0 and 1 milliampere) was not plotted due to the difficulty in measuring these very small currents. This portion of the curve rises slowly at first and then rapidly as shown in Figure 10-27. Your $V_F$ measurements are significant because they indicate the typical voltages you would expect across an LED operating within its normal range of $I_F$ values.

Each time you increased $I_F$ to a higher value, you should have noticed an increase in the intensity of the light produced by the LED. The radiant power produced by the LED increased linearly with current.

## Procedure (continued)

8. Remove the LED, and then, install it again in the opposite direction (reverse its leads). Adjust $R_1$ for minimum resistance to apply the full power supply voltage to the LED and its resistor. Then, turn on your Trainer.

9. Measure the voltage across the LED and determine if it is on (emitting light) or off (not emitting light). Turn off your Trainer.

## Discussion

In this final portion of the experiment, you reversed the LED's leads so that it was reverse-biased. Under these conditions, the voltage across the LED should have been equal to the input supply voltage. When reverse-biased, the LED acts like an open, or offers a very high reverse resistance, just like an ordinary PN junction diode. The LED should have been off (not emitting light) while it was reverse-biased.

# UNIT SUMMARY

Optoelectronic devices are widely used to perform various functions in electronic equipment. They may be divided into three general categories: light-sensitive devices, light-emitting devices, and light transmissive devices. The light sensitive category includes devices such as photoconductive cells, photovoltaic cells, photodiodes, and phototransistors. These devices respond to changes in light intensity by either generating an output voltage or by changing their resistance. The light-emitting category includes components such as incandescent and neon lamps, but these older components are no longer widely used. They have been replaced by light-emitting diodes or LEDs. LEDs can be designed to emit light over narrowly confined regions of the light spectrum when subjected to relatively low currents and voltages. Infrared emitting diodes are special purpose LEDs that emit energy in the infrared region of the spectrum. They have characteristics which make them highly compatible with other types of solid-state devices such as integrated circuits and transistors. They also provide many of the same benefits that are realized with most types of solid-state devices. The light transmissive or reflective category includes the liquid crystal display, or LCD. The LCD is very popular in applications where low power consumption is an important consideration. The true potential of these important components are just now being realized and continued growth in the optoelectronics field is almost certain.

There are many applications for optoelectronic devices. These components may be used individually or combined to perform many functions. The LED is often combined with a photodiode or phototransistor to provide optical coupling between circuits, or to detect the presence of an object. Individual LEDs are used as indicators and combinations of LEDs are used to display numbers. The number of possible applications are endless. The liquid-crystal display is another numeric display device that has applications similar to those for LED displays.

The LCD uses low voltage and low currents and can be easily read in bright light. In fact, the brighter the light the easier the display is to read. It also has the advantage of being visible from either side of the display. However, it requires a controlled beam to form the display.

# UNIT EXAMINATION

The following multiple choice examination is designed to test your understanding of the material presented in this unit. Place a check beside the multiple choice answer (A, B, C, or D) that you feel is most correct. When you have completed the examination, compare your answers to the correct ones that appear in the Examination Answer Sheet which follows.

1. Light is commonly measured by the
   A. photometric system.
   B. radiometric system.
   C. both the radiometric and photometric systems.
   D. isometric system.

2. The photometric system is used to measure
   A. visible light only.
   B. infrared light only.
   C. both visible and infrared light.
   D. the entire light spectrum.

3. Light is thought to consist of many tiny particles or bundles of energy called
   A. neutrons
   B. cosmic particles
   C. protons
   D. photons

4. The resistance of a photoconductive cell will be extremely high when
   A. it is exposed to a high intensity light.
   B. no light strikes its surface.
   C. it is reverse-biased.
   D. it is forward-biased.

5. The photovoltaic cell responds to changes in light intensity by
   A. varying its transconductance.
   B. changing its conductivity.
   C. changing its resistance.
   D. generating an output voltage.

6. The photodiode may be used in the
   A. photoconductive or photovoltaic mode.
   B. photovoltaic mode only.
   C. transconductive mode.
   D. photoconductive mode only.

7. When we compare photodiodes and phototransistors we find that phototransistors have a
   A. higher output current and a faster response.
   B. lower output current and a faster response.
   C. higher output current but a slower response.
   D. lower output current and a slower response.

8. The LED emits light when the holes and electrons near its junction.
   A. separate.
   B. recombine.
   C. are forced to flow away from the junction.
   D. are forced to flow toward the junction without recombining.

9. Under normal operating conditions the forward voltage across an LED would most likely be between
   A. 7 and 10 volts.
   B. 3 and 5 volts.
   C. 1 and 2 volts.
   D. 0 and 1 volt.

10. When we compare the LED with an incandescent lamp we find that the LED has a
    A. slower response and produces less light.
    B. faster response and produces more light.
    C. slower response but produces more light.
    D. faster response but produces less light.

11. When an LED and phototransistor are mounted in the same package so that they are connected by a light beam which can transmit electrical signals, the device formed is known as
    A. an optical coupler.
    B. an optical limit switch.
    C. a photo-darlington circuit.
    D. an LED display.

12. If an LED is connected as shown in Figure 10-31B and E equals 12 volts while $I_F$ must equal 20 milliamperes, the required value of $R_S$ would be (assume the diode voltage drop to be 1.6 volts)
    A. 595 ohms.
    B. 1040 ohms.
    C. 51.5 ohms.
    D. 515 ohms.

13. An advantage of LCD's over LED displays is
    A. LCD's emit infrared light.
    B. LCD's have lower power requirements.
    C. LCD's emit visible light.
    D. LCD's require an external light source.

14. Application of an electric field to a liquid crystal causes the material to
    A. change its optical characteristics.
    B. emit visible light.
    C. become transparent.
    D. emit infrared light.

# EXAMINATION ANSWERS

1. C — The radiometric and photometric systems are used but they are not identical.

2. A — The photometric system deals with visible light which can be perceived by the human eye.

3. D — Light contains photons or tiny particles of radiant energy which move through space in much the same way as radio waves.

4. B — When completely dark the cell's resistance is highest and the resistance decreases as light intensity increases.

5. D — The voltage generated by the photovoltaic cell increases with an increase in light intensity.

6. A — When used in the photoconductive mode the photodiode acts like a light-sensitive resistor but when used in the photovoltaic mode it generates an output voltage.

7. C — Phototransistors produce output currents in the milliampere range as compared to photodiodes which produce output in the microampere range. Unfortunately, the higher output is obtained by sacrificing the speed of response.

8. B — The recombining of electrons and holes results in the emission of photons.

9. C — A typical voltage drop for an LED is 1.6 volts but this value could be slightly higher or lower due to variations in LED characteristics.

10. D — LED's respond much faster than incandescent lamps but produce low output lightlevels.

11. A — Optical couplers are used to transmit electrical signals but provide a high degree of electrical isolation.

12. D — This problem would be solved as follows.

$$R_S = \frac{12 - 1.6}{.02} - 5\ \Omega$$

$$R_S = \frac{10.4}{.02} - 5\ \Omega$$

$$R_S = 520 - 5\ \Omega$$

$$R_S = 515 \text{ ohms}$$

13. B — LCD's have much lower operating voltage and current requirements than LED displays.

14. A — Application of an electrical field to a LCD causes the liquid crystal to change its optical characteristics from transparent to opaque.

# INDEX

A

Alloyed method, 2-31
Alpha (α), 6-10
AND gate IC, 9-34
Angstrom (Å), 10-7, 10-45
Atom, 1-12

B

Barrier voltage, 2-8, 2-17, 2-19
Beta cutoff frequency, 6-21
Bidirectional trigger diodes, 8-31
Bidirectional triode thyristors, 8-24
Bipolar
  Ic, 9-14
  transistor,5-3

C

CMOS, 7-3, 7-29, 9-39
Collector-to-Base Leakage current, 6-12
Collector-to-emitter Leakage current, 6-22
Combinational logic circuit, 9, 37
Common-base circuit, 5-20, 6-6
Common-collector circuit, 5-23, 6-31
Common-drain circuits, 7-39
Common-emitter circuit, 5-21, 6-16
Common-gate circuits, 7-37
Common-source circuits, 7-35
Covalent bond, 1-14
Current gain, 6-10, 6-19, 6-32

D

Depletion-mode devices, 7-22
Derating factor, 3-9
Diffused junction, 2-31
Diffusion, 9-14
Digital IC, 9-34
Diode Applications, 2-38
Diode biasing, 2-10
Diode characteristics, 2-16
Diode construction, 2-30
Diode packaging, 2-32
Diode symbols, 2-21
Doping, 1-21

E

Emitter current, 5-3
emitter follower, 5-23, 6-31
Enhancement-mode devices, 7-26

F

FET, 7-3, 7-35
Flat-pack, 9-29
Forward-bias, 2-9
Forward breakover Voltage, 8-9

G

Gate-to-Source cutoff, 7-7
Germanium (GE), 1-10
Germanium diode, 2-11
Grown Junctions, 2-30
Gunn-effect diodes, 4-23

H

Hole, 1-18
High Frequency diodes, 4-21
Hybird IC, 9-3, 9-24

I

IGFET, 7-3, 7-22
IMPATT diode, 4-25
intrinsic material, 1-15
Integrated circuits, 1-6, 9-3
Ions, 2-6

J

JFET, 7-3, 7-5
Junction diodes, 2-7

**K**

Knee, 3-7

**L**

LCD, 10-51
LED, 10-39
Leakage Current, 2-13, 2-21
Light emitting diode, 10-3
Light Sensitive Devices, 10-16
Linear IC, 9-41
Liquid crystal displays,
Lumen, 10-12
Luminance, 10-13
Luminous
    flux, 10-12
    power, 10-12

**M**

Maximum zener current($I_{ZM}$), 3-11
mobile charges, 2-6
Monolithic IC, 9-12
MOSFET, 7-3, 7-22, 9-39
MOS IC, 9-19

**N**

NAND, 9-38
Negative
    conductance, 4-7
    resistance, 4-6
NOR, 9-40
NOT gate, 9-34, 9-37
NPN configuration, 5-5
N-Type semiconductors, 1-21, 2-6

**O**

operational amplifier, 9-41
optical coupler, 10-49
Optoelectronics devices, 10-3
OR gate, 9-35

**P**

Peak inverse voltage (PIV), 2-20
Pentavalent material, 1-21
Photoconductive cells, 10-16
Photodiodes, 10-22
Photometeric system, 10-9, 10-11
Phototransistors, 10-26
Photovoltaic cells, 10-19
Pinch-off voltage, 7-8
PIN diode, 4-21
PN Junction, 2-6
PNP configuration, 5-5
Power temperature curve, 3-8
P-Type semiconductors, 1-22, 2-6
PUT, 8-44

**Q**

quantum efficiency, 10-24
quantum theory, 10-8

**R**

Radiance, 10-10, 10-13
Radiant
    flux, 10-10
    energy (Qe), 10-10
    power, 10-10, 10-44
Radio waves, 10-6
Radiometric system, 10-9, 10-10
Rectification, 2-38
Reverse bias, 2-12

**S**

Schottky diodes, 4-22
SCR, 8-6
Scotopic, 10-11
Semiconductors, 1-5
Seven-segment displays (7-segment), 10-48
Silicon controlled rectifiers, 8-6
Silicon (si), 1-10
Silicon diode, 2-11
Solar Cell, 10-19
Solid-state components, 1-5
SMT, 9-30

T

Testing
    bipolar transistors, 5-27
    NPN transistors, 5-28
    PNP transistors, 5-28
Thin-film IC, 9-3, 9-21
Thick-film IC, 9-3, 9-23
Thyristor, 8-3
Transconductance, 7-10
Transistor Amplifier, 5-16
Transferred Electron effect, 4-23
Trivalent Material, 1-21
Tunnel diode, 4-5

U

UJT, 8-36
Ultraviolet ray, 10-8

V

VI curve, 2-16, 3-6, 4-6, 8-9, 8-26
Valley
    current 4-3, 4-6
    voltage, 4-3, 4-6
Varactor diode, 4-3, 4-12

W

Wavelength, 10-6, 10-45

Z

Zener
    current ($I_Z$), 3-6
    diodes, 2-18, 3-5
    knee impedance ($Z_{ZK}$), 3-14
    test current ($I_{ZT}$), 3-7
    impedance ($Z_{ZT}$), 3-14
    region, 3-6
    regulator circuit, 3-16
    voltage ($V_Z$), 3-7
    voltage temperature coefficient, 3-11